Auditing Rational Adversaries to Provably Manage Risks

Jeremiah Blocki, Nicolas Christin, Anupam Datta, and Arunesh Sinha

May 23, 2012

CMU-CyLab-12-011

CyLab
Carnegie Mellon University
Pittsburgh, PA 15213

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
23 MAY 2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Auditing Rational Adversaries to Provably Manage Risks	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Carnegie Mellon University,CyLab,Pittsburgh,PA,15213**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Audits to detect policy violations coupled with punishments are essential to manage risks stemming from inappropriate information use by authorized insiders in organizations that handle large volumes of personal information (e.g., in healthcare, finance, Web services sectors). Our main result is an audit mechanism that effectively manages organizational risks by balancing the cost of audit and punishment against the expected loss from policy violations. We model the interaction between an organization (defender) and an employee (adversary) as a suitable repeated game. We assume that the defender is fully rational and the adversary is near-rational (i.e., acts rationally with high probability and in a byzantine manner otherwise). The mechanism prescribes a strategy for the defender that when paired with the adversary?s best response to it yields an asymmetric subgame perfect equilibrium. This equilibrium concept, which we define, implies that the defender?s strategy is approximately optimal (she might only gain a small bounded amount of utility by deviating) while the adversary does not gain at all from deviating from her best response strategy. We provide evidence that a number of parameters in the game model can be estimated from prior empirical studies, suggest specific studies that can help estimate other parameters, and design a learning algorithm that the defender can use to provably learn the adversary?s private incentives. Finally, we use our model to predict observed practices in industry (e.g., differences in punishment rates of doctors and nurses for the same violation) and the effectiveness of policy interventions (e.g., data breach notification laws and government audits) in encouraging organizations to conduct more thorough audits.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**28**	

Auditing Rational Adversaries to Provably Manage Risks*

Jeremiah Blocki Nicolas Christin Anupam Datta Arunesh Sinha
Carnegie Mellon University, Pittsburgh, PA
{*jblocki, nicolasc, danupam, aruneshs*}*@cmu.edu*

May 22, 2012

Abstract

Audits to detect policy violations coupled with punishments are essential to manage risks stemming from inappropriate information use by authorized insiders in organizations that handle large volumes of personal information (e.g., in healthcare, finance, Web services sectors). Our main result is an audit mechanism that effectively manages organizational risks by balancing the cost of audit and punishment against the expected loss from policy violations. We model the interaction between an organization (defender) and an employee (adversary) as a suitable repeated game. We assume that the defender is fully rational and the adversary is *near-rational* (i.e., acts rationally with high probability and in a byzantine manner otherwise). The mechanism prescribes a strategy for the defender that when paired with the adversary's best response to it yields an *asymmetric subgame perfect equilibrium*. This equilibrium concept, which we define, implies that the defender's strategy is approximately optimal (she might only gain a small bounded amount of utility by deviating) while the adversary does not gain at all from deviating from her best response strategy. We provide evidence that a number of parameters in the game model can be estimated from prior empirical studies, suggest specific studies that can help estimate other parameters, and design a learning algorithm that the defender can use to provably learn the adversary's private incentives. Finally, we use our model to predict observed practices in industry (e.g., differences in punishment rates of doctors and nurses for the same violation) and the effectiveness of policy interventions (e.g., data breach notification laws and government audits) in encouraging organizations to conduct more thorough audits.

1 Introduction

The importance of audit and accountability mechanisms to detect policy violations and punish violators has been recognized in computer security [35] as well as in recent public policy discussions on privacy protection [14, 53]. Specifically, experts from privacy enforcement agencies, industry, civil society and academia have recently developed a series of white papers on accountability-based privacy governance in which one recommendation is that organisations should have in place policies and procedures for enforcement of internal data protection rules and personnel who disregard those rules or misappropriate or misuse data should be subject to sanctions, including dismissal [14]. Indeed such violations and sanctions are routinely reported in the healthcare sector [1, 28, 33, 42] and we are beginning to see the emergence of commercial audit tools to assist in the process of detecting violations [19].

The central scientific question that this state of affairs raises is how to design effective audit and punishment schemes. This paper articulates a desirable property and presents an audit mechanism that provably achieves that property against a class of adversaries. The high-level observation here is that audits coupled with punishments

*This work was partially supported by the U.S. Army Research Office contract "Perpetually Available and Secure Information Systems" (DAAD19-02-1-0389) to Carnegie Mellon CyLab, the NSF Science and Technology Center TRUST, the NSF CyberTrust grant "Privacy, Compliance and Information Risk in Complex Organizational Processes," the AFOSR MURI "Collaborative Policies and Assured Information Sharing," and HHS Grant no. HHS 90TR0003/01. Jeremiah Blocki was also partially supported by a NSF Graduate Fellowship. Arunesh Sinha was also partially supported by the CMU CIT Bertucci Fellowship. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

constitute a mechanism for managing risks—a suitable audit mechanism effectively manages organizational risks by balancing the cost of audit and punishment against the expected loss from policy violations.

At a technical level, we model the interaction between the organization (the defender) and its employee (the adversary) as a repeated extensive form game with imperfect information (the adversary's actions are not observable to the defender) and public signals (the outcome of the audit, i.e. how many violations are detected and the rate of inspection and punishment are revealed publicly). The game model (described in Section 3) augments the model in previous work [10] by incorporating the incentives of rational adversaries. Adversaries benefit from violations they commit (e.g., by selling personal data) and suffer due to punishments imposed for detected violations. We refer to the benefit derived by the adversary from committing violations as her *personal benefit* and assume that it is initially hidden from the defender. In order to account for adversaries who are generally rational but may sometimes act irrationally, we consider *near-rational* adversaries who choose actions that maximize their (expected) utility with probability $(1-\epsilon)$ and with probability ϵ act arbitrarily. This model is inspired by the *trembling hand* assumption from game theory [21]. The model also includes a loss for the organization if the punishment rate is set too high (to capture loss in productivity resulting, e.g., from employee dismissal, rehiring and training [2, 19]), in addition to the cost of inspecting and the loss due to policy violations. The model generalizes from the situation in which the defender interacts with a single adversary to one where she interacts with multiple, non-colluding adversaries via a natural product game construction that we define.

Our main contribution is an audit and accountability mechanism that proceeds in two phases for each audit cycle—a *detection and estimation phase* and an *audit phase*. For the audit phase, we design a strategy for the defender such that the adversary's best response to it yields an *asymmetric subgame perfect equilibrium*. This equilibrium concept, which we define, implies that the defender's strategy is approximately optimal (she might only gain a small bounded amount of utility by deviating) while the adversary does not gain at all from deviating from her best response strategy (see Section 4). We define this equilibrium concept by adapting the standard notion of approximate subgame perfect equilibrium, which has a symmetric flavor and permits both players to obtain small gains by unilaterally deviating from their equilibrium strategy. We believe that the symmetric equilibrium concept is unsuitable for our security application where an adversary who deviates motivated by a small gain could cause a big loss for the organization. This asymmetry is also indicative of the nature of the audit game—it is not a game between peers, but one in which the defender has greater power since she gets to decide the rate of inspection and the punishment level.

For the detection and estimation phase, we rely on standard techniques used in risk assessment [2,23,30,40,44,49] to estimate the parameters of the game model. We provide evidence that a number of parameters in the game model can be estimated from prior studies [17,46,48,57] and suggest specific studies that can help estimate other parameters. In addition, we design a learning algorithm (described in Section 5) that the defender can use to provably learn the adversary's personal benefit parameter. The learning algorithm works in an *adversarial labeling of training data points* setting [3, 32]. In general, it is impossible to guarantee any learning if the adversary can label training data points arbitrarily, thus, some constraints have to be imposed on the adversary for such learning to work [3, 32]. We use a novel game-theoretic argument to impose such constraints, as the impatient near rational adversary acts in a manner that maximizes her immediate utility. The defender interacts with the adversary iteratively adjusting the rate of inspection and the level of punishment to provably learn the adversary's personal benefit parameter. This technique is quite general and can be used in other adversarial machine learning settings.

Finally, we use our model to predict observed practices in industry (e.g., differences in punishment rates of doctors and nurses for the same violation) and the effectiveness of policy interventions (e.g., data breach notification laws and government audits) in encouraging organizations to conduct more thorough audits (see Section 6). We present comparisons to additional related work in Section 7 and conclusions and directions for future work in Section 8.

2 Overview

In this section, we provide an overview of our model using a motivating scenario that will serve as a running example for this paper. Consider a "Hospital X" with employees in different roles (doctors, nurses, accountants). X has an internal policy that mandates weekly HIPAA compliance audits, notably to ensure that accesses to personal health records are legitimate. Given budget constraints, X cannot check every single access. The first step in the audit

process is to analyze the access logs using an automated tool that ranks accesses by their probability of being a violation. Hospital X assesses the (monetary) impact of different types of violations and decides what subset to focus on by balancing the cost of audit and the expected impact ("risk") from policy violations. This type of audit mechanism is common in practice [30, 40, 44, 49].

We provide a game model for this audit process that incorporates behavioral factors in risk assessment. We assume that employees are rational: while they are not trying to disrupt their organization's business, they may violate certain policies if they benefit from doing so. The organization (e.g., Hospital X) is a rational entity that is trying to maximize its expected utility, i.e., balance audit costs against risks from employee non-compliance.

More precisely, an employee ("adversary," $\mathcal{A}$) executes tasks, i.e., actions that are permitted as part of their job. We only consider tasks that can later be audited, e.g., through inspection of logs. For example, in X the tasks are accesses to health records. We can distinguish $\mathcal{A}$'s tasks between legitimate tasks and violations of a policy. Different *types of violations* may have different impact on the organization. We assume that there are K different types of violations that $\mathcal{A}$ can commit. The economic impact of a violation on the organization depends on its type. Examples of violations of different types in Hospital X include inappropriate access to a celebrity's health record, or access to a health record leading to identity theft. $\mathcal{A}$ benefits by committing violations: the benefit is quantifiable using information from existing studies, human judgment, or by the algorithmic technique we propose in Section 5. For example, reports [2, 17] indicate that on average the *personal benefit* of a hospital employee from selling a common person's health record is $\$50$. On the other hand, if $\mathcal{A}$ is caught committing a violation then she is punished according to the *punishment policy* used by $\mathcal{D}$. In the case of Hospital X, employees could be terminated, as happened in similar recent incidents [33, 42].

The organization $\mathcal{D}$ can classify each adversary's task by type. However, $\mathcal{D}$ cannot determine with certainty whether a particular task is legitimate or a violation without investigating. Furthermore, $\mathcal{D}$ cannot inspect all of $\mathcal{A}$'s tasks due to *budgetary constraints*. As such, some violations may go undetected *internally*, but could be detected *externally*. Governmental audits, whistle-blowing, patient complains [45, 57] are all examples of situations that could lead to external detection of violations. Externally detected violations usually cause more economic damage than internally caught violations. For instance, as indicated in the 2011 Ponemon Institute report [48], a patient whose privacy has been violated is probably more likely to leave (and possibly sue) a hospital if they discover the violation on their own than if the hospital detects the violation and proactively notifies the patient.

The economic impact of a violation is a combination of *direct and indirect costs*; direct costs include breach notification and remedial cost, and indirect costs include loss of customers and brand value. For example, the 2010 Ponemon Institute report [46] states that the average cost of privacy breach *per record* in health care is $\$301$ with indirect costs corresponding to about two thirds of that amount. Of course, certain violations may result in much higher direct costs, e.g., $\$25,000$ per record (up to $\$250,000$ in total) in fines alone in the state of California [42]. While these amounts may incentivize organizations to use aggressive audits, they have to be balanced with the fact that severe punishment policies result in a hostile work environment, low employee motivation and failure to attract new talent — causing economic losses for the organization [12].

In other words, the organization needs to balance auditing costs, potential economic damages due to violations and the economic impact of the punishment policy. The employees need to weigh their gain from violating policies against loss from getting caught by an audit and punished. The actions of one party impact the actions of the other party: if employees never violate, the organization does not need to audit; likewise, if the organization never audits, employees can violate policies in total impunity. Given this strategic interdependency, we model the auditing process as a *repeated game* between the organization and its employees, where the game repeats over discrete rounds characterizing audit cycles. The game is parameterized by quantifiable variables such as the personal benefit of employee, the cost of breach, and the cost of auditing, among others.

3 Model

We begin by providing a high level view of the audit process (Section 3.1), before describing the audit game in detail (Section 3.2). Finnaly, we describe estimation and detection of parameters of the audit game (Section 3.3).

3.1 The Audit Process

In practice, the organization is not playing a repeated audit game against a specific employee, but against all of its n employees at the same time. However, if we assume that 1) a given employee's actions for a type of task are independent of her actions for other types, and that 2) employees do not collude with other employees and act independently, we can decompose the overall game into $\mathsf{n}K$ independent *base* repeated games, that the organization plays in parallel. One base repeated game corresponds to a given type of access k by a given employee $\mathcal{A}$, and will be denoted by $\mathcal{G}_{\mathcal{A},k}$. Each game $\mathcal{G}_{\mathcal{A},k}$ is described using many parameters like *loss due to violations*, *personal benefit* for employee, etc. We abuse notation in using $\mathcal{G}_{\mathcal{A},k}$ to refer to a base repeated game of type k with any value of the parameters.

In our proposed audit process the organization follows the steps listed below in each audit cycle for every game $\mathcal{G}_{\mathcal{A},k}$. Assume the parameters of the game have been estimated and the equilibrium audit strategy computed for the first time auditing is performed.

before audit:
1. If any parameter changes go to step 2 else go to *audit*.
2. Estimate parameters. Compute equilibrium of $\mathcal{G}_{\mathcal{A},k}$.
audit:
3. Audit using actions of the computed equilibrium.

Note that the parameters of $\mathcal{G}_{\mathcal{A},k}$ may change for any given round of the game, resulting in a different game. However, neither $\mathcal{D}$ nor $\mathcal{A}$ knows when that will happen. As such, since the horizon of $\mathcal{G}_{\mathcal{A},k}$ with a fixed set of parameters is infinite, we can describe the interaction between the organization and its employees with an infinitely repeated game for the period in which the parameters are unchanged (see [21] for details). Thus, the game $\mathcal{G}_{\mathcal{A},k}$ is an infinitely repeated game of *imperfect information* since $\mathcal{A}$'s action is not directly observed. Instead, noisy information about the action, called a *public signal* is observed. The public signal here consists of a) the detected violations b) number of tasks by $\mathcal{A}$ and c) $\mathcal{D}$'s action. The K parallel games played between $\mathcal{A}$ and $\mathcal{D}$ can be composed in a natural manner into one repeated game (which we call $\mathcal{G}_{\mathcal{A}}$) by taking the product of action spaces and adding up utilities from the games.

3.2 Formal Description

In the remainder of this section, we focus on the base repeated games $\mathcal{G}_{\mathcal{A},k}$. We use the following notations in this paper:

- Vectors are represented with an arrow on top, e.g., $\vec{v}$ is a vector. The i^{th} component of a vector is given by $\vec{v}\,(i)$. $\vec{v} \leq \vec{a}$ means that both vectors have the same number of components and for any component i, $\vec{v}\,(i) \leq \vec{a}\,(i)$.
- Random variables are represented in boldface, e.g., $\mathbf{x}$ and $\mathbf{X}$ are random variables.
- $E(\mathbf{X})[q, r]$ denotes the expected value of random variable X, when particular parameters of the probability mass function of $\mathbf{X}$ are set to q and r.
- We will use a shorthand form by dropping $\mathcal{A}$, k and the vector notation, as we assume these are implicitly understood for the game $\mathcal{G}_{\mathcal{A},k}$. That is, a quantity $\vec{x}_{\mathcal{A}}(k)$ will be simply denoted as x. We use this form whenever the context is restricted to game $\mathcal{G}_{\mathcal{A},k}$ only.

$\mathcal{G}_{\mathcal{A},k}$ is fully defined by the players, the time granularity at which the game is played, the actions the players can take, and the utility the players obtain as a result of the actions they take. We next discuss these different concepts in turn.

Players: The game $\mathcal{G}_{\mathcal{A},k}$ is played between the organization $\mathcal{D}$ and an adversary $\mathcal{A}$. For instance, the players are hospital X and a nurse in X.

Round of play: In practice, audits for all employees and all types of access are performed together and usually periodically. Thus, we adopt a discrete-time model, where time points are associated with rounds. Each round of play corresponds to an audit cycle. We group together all of the adversary's actions (tasks of a given type) in a given round. All games $\mathcal{G}_{\mathcal{A},k}$ are synchronized, that is all given rounds t in all games are played simultaneously.

Adversary action space: In each round, the adversary $\mathcal{A}$ chooses two quantities of type k: the number of tasks she performs, and the number of such tasks that are violations. If we denote by U_k the maximum number of type k tasks that any employee can perform, then $\mathcal{A}$'s entire action space for $\mathcal{G}_{\mathcal{A},k}$ is given by $A_k \times V_k$ with $A_k = \{u_k, \ldots, U_k\}$

($u_k \leq U_k$) and $A_k = \{1, \ldots, U_k\}$. Let $\vec{a}^t_{\mathcal{A}}$ and $\vec{v}^t_{\mathcal{A}}$ be vectors of length K such that the components of vector $\vec{a}$ are the number of tasks of each type that $\mathcal{A}$ performs at time t, and the components of vector $\vec{v}$ are the number of violations of each type. Since violations are a subset of all tasks, we always have $\vec{v}^t_{\mathcal{A}} \leq \vec{a}^t_{\mathcal{A}}$. In a given audit cycle, $\mathcal{A}$'s action in the game $\mathcal{G}_{\mathcal{A},k}$ is defined by $\langle \vec{a}^t_{\mathcal{A}}(k), \vec{v}^t_{\mathcal{A}}(k) \rangle$, that is $\langle a^t, v^t \rangle$ in shorthand form, with $a^t \in A_k$ and $v^t \in A_k$.

Instead of being perfectly rational, we model $\mathcal{A}$ as playing with a *trembling hand* [21]. Whenever $\mathcal{A}$ chooses to commit v^t violations in as given round t, she does so with probability $1 - \epsilon_{th}$, but, with (small) probability ϵ_{th} she commits some other number of violations sampled from an unknown distribution D^t_0 over all possible violations. In other words, we allow $\mathcal{A}$ to act completely arbitrarily when she makes a mistake. For instance, a nurse in X may lose her laptop containing health records leading to a breach.

Defender action space: $\mathcal{D}$ also chooses two quantities of type k in each round: the number of inspections to perform, and the punishment to levy for each type-k violation detected. Let $\vec{s}^t_{\mathcal{A}}$ be the vector of length K such that components of vector $\vec{s}^t_{\mathcal{A}}$ are the number of inspections of each type that $\mathcal{D}$ performs in round t. The number of inspections that $\mathcal{D}$ can conduct is bounded by the number of tasks that $\mathcal{A}$ performs, and thus, $\vec{s}^t_{\mathcal{A}} \leq \vec{a}^t_{\mathcal{A}}$. $\mathcal{D}$ uses a log analysis tool $\mathcal{M}$ to sort accesses according to the probability of them being a violation. Then, $\mathcal{D}$ chooses the top $\vec{s}^t_{\mathcal{A}}(k) = s^t$ tasks from the sorted output of $\mathcal{M}$ to inspect in game $\mathcal{G}_{\mathcal{A},k}$. Inspection is assumed perfect, i.e., if a violation is inspected, it is detected. The number of inspections is bounded by budgetary constraints. Denoting the function that outputs cost of inspection for each type of violation by $\vec{C}$, we have $\vec{C}(k)(s^t) \leq \vec{b}^t(k)$ where $\vec{b}^t(k)$ defines a per-employee, per-type budget constraint. The budget allocation problem is an optimization problem depending on the audit strategy. We present this problem assuming our proposed audit strategy in Appendix B.3.

$\mathcal{D}$ also chooses a punishment rate $\vec{P}^t_{\mathcal{A}}(k) = P^t$ (fine per violation of type k) in each round t to punish $\mathcal{A}$ if violations of type k are detected. The punishment rate P^t is bounded by a maximum punishment P_f corresponding to the employee being fired, and the game terminated.

Finally, $\mathcal{D}$'s choice of the inspection action can depend only on $\mathcal{A}$'s total number of tasks, since the number of violations is not observed. Thus, $\mathcal{D}$ can choose its strategy as a function from number of tasks to inspections and punishment even before $\mathcal{A}$ performs its action. In fact, we simulate $\mathcal{D}$ acting first and the actions are *observable* by requiring $\mathcal{D}$ to commit to a strategy and provide a proof of honoring the commitment. Specifically, $\mathcal{D}$ computes its strategy, makes it public and provides a proof of following the strategy after auditing is done. The proof can be provided by maintaining an audit trail of the audit process itself.

Outcomes: We define the outcome of a single round of $\mathcal{G}_{\mathcal{A},k}$ as the number of violations detected in internal audit and the number of violations detected externally. We assume that there is a fixed exogenous probability p ($0 < p < 1$) of an internally undetected violation getting caught externally. Due to the probabilistic nature of all quantities, the outcome is a random variable. Let $\vec{\mathbf{O}}^t_{\mathcal{A}}$ be the vector of length K such that the $\vec{\mathbf{O}}^t_{\mathcal{A}}(k) = \mathbf{O}^t$ represents the outcome for the t^{th} round for the game $\mathcal{G}_{\mathcal{A},k}$. Then $\mathbf{O}^t$ is a tuple $\langle \mathbf{O}^t_{int}, \mathbf{O}^t_{ext} \rangle$ of violations caught internally and externally. As stated earlier, we assume the use of a log analysis tool $\mathcal{M}$ to rank the accesses with more likely violations being ranked higher. Then, the probability mass function for $\vec{\mathbf{O}}^t_{int}$ is a distribution parameterized by $\langle a^t, v^t \rangle$, s and $\mathcal{M}$. The worst performance of $\mathcal{M}$ is when the s accesses to be inspected are chosen at random, resulting in a hyper-geometric distribution with mean $v^t \alpha^t$, where $\alpha^t = s^t / a^t$. We assume that the mean of the distribution is $\mu(\alpha^t) v^t \alpha^t$, where $\mu(\alpha^t)$ is a function dependent on α^t that measures the *performance* of $\mathcal{M}$ and $\forall \alpha^t \in [0, 1]. \mu \geq \mu(\alpha^t) \geq 1$ for some constant μ (μ is overloaded here). Note that we must have $\mu(\alpha^t)\alpha^t \leq 1$, and further, we assume that $\mu(\alpha^t)$ is monotonically non-increasing in α^t. The probability mass function for $\mathbf{O}^t_{ext}$ conditioned on $\mathbf{O}^t_{int}$ is a binomial distribution parameterized by p.

Utility functions: In a public signaling game like $\mathcal{G}_{\mathcal{A},k}$, the utilities of the players depend only on the public signal and their own action, while the strategies they choose depend on the history of public signals [37]. The utility of the repeated game is defined as a (delta-discounted) sum of the expected utilities received in each round, where the expectation is taken with respect to the distribution over histories. Let the discount factor for $\mathcal{D}$ be $\delta_{\mathcal{D}}$ and for any employee $\mathcal{A}$ be $\delta_{\mathcal{A}}$. We assume that $\mathcal{D}$ is patient, i.e., future rewards are almost as important as immediate rewards, and $\delta_{\mathcal{D}}$ is close to 1. $\mathcal{A}$ is less patient than $\mathcal{D}$ and hence $\delta_{\mathcal{A}} < \delta_{\mathcal{D}}$.

Defender utility function: $\mathcal{D}$'s utility in a round of the game $\mathcal{G}_{\mathcal{A},k}$ consists of the sum of the *cost of inspecting* $\mathcal{A}$'s actions, the monetary *loss from a high punishment rate* for $\mathcal{A}$, and *direct and indirect costs* of violations. In essence, $\mathcal{D}$ has to find the right balance between inspecting with higher coverage (which incurs high costs), letting violations occur (which results in direct/indirect costs) and stifling employee productivity by setting a high punishment rate.

As discussed before, inspection costs are given by $C(s^t)$ where $C = \vec{C}(k)$ is a function denoting the cost of inspecting type-k tasks. Similarly, the monetary loss from losing employee's productivity due to fear of punishment is given by $e(P^t)$, where $e = \vec{e}_A(k)$ is a function for type-k tasks. The functions in $\vec{C}$ and $\vec{e}$ must satisfy the following constraints: 1) they should be monotonically increasing in the argument and 2) $\vec{C}(k) \geq 0$, $\vec{e}_A(k) \geq 0$ for all k.

We characterize the effect of violations on the organization's indirect cost similarly to the reputation loss as in previous work [10]. Additionally, the generic function described below is capable of capturing direct costs, as shown in the example following the function specification. Specifically, we define a function r_k (r in shorthand form) that, at time t, takes as input the number of type-k violations caught internally, the number of type-k violations caught externally, and a time horizon τ, and outputs the overall loss at time $t + \tau$ due to these violations at time t. r is stationary (i.e., independent of t), and externally caught violations have a stronger impact on r than internally detected violations. Further, $r(\langle 0, 0\rangle, \tau) = 0$ for any τ (undetected violations have 0 cost), and r is monotonically decreasing in τ and becomes equal to zero for $\tau \geq m$ (violations are forgotten after a finite amount of rounds). As in previous work [10], we construct the utility function at round t by immediately accounting for future losses due to violations occurring at time t. This allows us to use standard game-theory results, while at the same time, providing a close approximation of the defender's loss [10]. With these notations, $\mathcal{D}$'s utility at time t in $\mathcal{G}_{\mathcal{A},k}$ is

$$\mathbf{Rew}^t_{\mathcal{D}}(\langle s^t, P^t\rangle, \mathbf{O}^t) = -\sum_{j=0}^{m-1} \delta^j_{\mathcal{D}} r(\mathbf{O}^t, j) - C(s^t) - e(P^t)\,. \tag{1}$$

This per-round utility is always negative (or at most zero). As is typical of security games (e.g., [24, 55] and related work), implementing security measures does not provide direct benefits to the defender, but is necessary to pare possible losses. Hence, the goal for the defender is to have this utility as close to zero as possible.

The above function can capture direct costs of violations as an additive term at time $\tau = 0$. As a simple example [10], assuming the average direct costs for internally and externally caught violations are given by R^D_{int} and R^D_{ext}, and the function r is linear in the random variables $\vec{\mathbf{O}}^t_{int}$ and $\vec{\mathbf{O}}^t_{ext}$, r can be given by

$$r(\mathbf{O}^t, \tau) = \begin{cases} (c + R^D_{int})\mathbf{O}^t_{int} + (\psi c + R^D_{ext})\mathbf{O}^t_{ext} & \text{for } \tau = 0 \\ \delta^\tau c(\mathbf{O}^t_{int} + \psi \cdot \mathbf{O}^t_{ext}) & \text{for } 1 \leq \tau < m \\ 0 & \text{for } \tau \geq m, \end{cases}$$

where $\delta \in (0, 1)$ and $\psi \geq 1$. Then Eqn. (1) reduces to

$$\mathbf{Rew}^t_{\mathcal{D}}(\langle s^t, P^t\rangle, \mathbf{O}^t) = -R_{int}\mathbf{O}^t_{int} - R_{ext}\mathbf{O}^t_{ext} - C(s^t) - e(P^t)\,, \tag{2}$$

with $R_{int} = R^I_{int} + R^D_{int}$, $R^I_{int} = c(1 - \delta^m\delta^m_{\mathcal{D}})/(1 - \delta\delta_{\mathcal{D}})$ and $R_{ext} = \psi R^I_{int} + R^D_{ext}$.

Adversary utility function: We define $\mathcal{A}$'s utility as the sum of $\mathcal{A}$'s *personal benefit* gained by committing violations and the *punishment* that results due to detected violations. Personal benefit is a monetary measure of the benefit that an employee gets out of violations. It includes all kinds of benefits, e.g., curiosity, actual monetary benefit (by selling private data), revenge, etc. It is natural that the personal benefit of an employee is only known to that employee. Our model of personal benefit of an employee $\mathcal{A}$ is linear and is defined by a rate of personal benefit for each type of violation given by the vector $\vec{I}_{\mathcal{A}}$ of length K. Further, we assume that the upper bounds on the private benefit are publicly known and given by $\vec{I}_{\max,\mathcal{A}}$. The punishment is the vector $\vec{P}^t_{\mathcal{A}}$ of length K chosen by $\mathcal{D}$, as discussed above. Using shorthand notation, $\mathcal{A}$'s utility, for the game $\mathcal{G}_{\mathcal{A},k}$, is:

$$\mathbf{Rew}^t_{\mathcal{A}}(\langle a^t, v^t\rangle, \langle s^t, P^t\rangle, \mathbf{O}^t) = Iv^t - P^t\left(\mathbf{O}^t_{int} + \mathbf{O}^t_{ext}\right)\,.$$

Observe that the utility function of a player depends on the public signal (observed violations, $\mathcal{D}$'s action) and the action of the player, which conforms to the definition of a repeated game with imperfect information and public signaling. In such games, the *expected* utility is used in computing equilibria.

Let $\alpha^t = s^t/a^t$ and $\nu(\alpha^t) = \mu(\alpha^t)\alpha^t$. Then, $E(\mathbf{O}^t_{int}) = \nu(\alpha^t)v^t$, and $E(\mathbf{O}^t_{ext}) = pv^t(1 - \nu(\alpha^t))$. The expected utilities in each round then become:

$$\begin{aligned} E(\mathbf{Rew}^t_{\mathcal{D}}) &= -\textstyle\sum_{j=0}^{m-1} \delta^j_{\mathcal{D}} E(r(\mathbf{O}^t, j))[v^t, a^t, \alpha^t] \\ &\quad -C(\alpha^t a^t) - e(P^t)\,, \\ E(\mathbf{Rew}^t_{\mathcal{A}}) &= Iv^t - P^t v^t\left(\nu(\alpha^t) + p(1 - \nu(\alpha^t))\right)\,. \end{aligned}$$

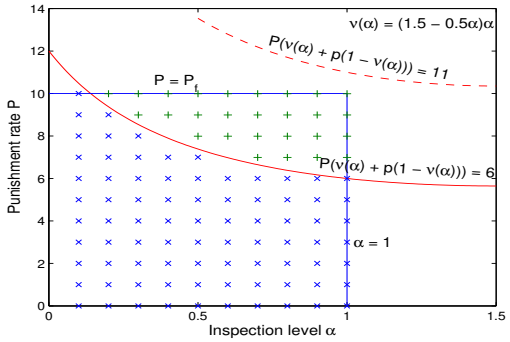

Figure 1: Feasible audit space: non-deterred ($\times$) and deterred ($+$) region for $I = \$6$. For $I = \$11$ deterred region is empty.

The expected utility of $\mathcal{A}$ depends only on the level of inspection and not on the actual number of inspections. For the example loss function given by Eqn. (2), the utility function of $\mathcal{D}$ becomes:

$$E(\mathbf{Rew}^t_{\mathcal{D}}) = -v^t(R_{int}\nu(\alpha^t) + R_{ext}p(1-\nu(\alpha^t))) - C(\alpha^t a^t) - e(P^t)\ .$$

In addition to the action dependent utilities above, the players also receive a fixed utility every round, which is the salary for $\mathcal{A}$ and value generated by $\mathcal{A}$ for $\mathcal{D}$. P_f depends on these values, and is calculated in Appendix B.2. Finally, the model parameters that may change over time are R_{ext}, R_{int}, p, function C, function e, function μ and I.
Graphical representation: A graphical representation of the utilities helps illustrate the ideas presented in the next two sections. (See Figure 1). Consider the 2-dimensional plane $R^{\alpha,P}$ spanned by α^t and P^t. We define a feasible audit space in $R^{\alpha,P}$ given by $0 \leq \alpha^t \leq 1$ and $0 \leq P^t \leq P_f$. $\mathcal{D}$'s actions are points in the feasible region. The expected utility of the adversary in each round is given by $v^t(I - P^t(\nu(\alpha^t) + p(1-\nu(\alpha^t))))$. Thus, the curve in $R^{\alpha,P}$ given by $I = P^t(\nu(\alpha^t) + p(1-\nu(\alpha^t)))$ is the separator between positive and negative expected utility regions for the adversary in each round. We call the region of positive expected utility inside the feasible region the *non-deterred region* and the region of negative utility inside the feasible region the *deterred region*.

3.3 Estimation and Detection

In this sub-section, we demonstrate that the parameters of the audit game can be estimated. We describe techniques of estimating and detecting changes in parameters of game $\mathcal{G}_{\mathcal{A},k}$, obtaining sample estimates in the process. Before getting to constant values, we state the functions that we use as concrete instances for the examples in this paper. We use simple linear functions for audit cost ($C(\alpha a) = C\alpha a$) and for punishment loss ($e(P) = eP$). The performance of $\mathcal{M}$ is dependent on the tool being used and we use a linear function for $\mu(.)$ to get $\nu(\alpha) = \mu\alpha - (\mu - 1)\alpha^2$, where μ is a constant. Further, we use the example loss function (with R_{int} and R_{ext}) stated in the last sub-section. We note that our theorems work with any function; these functions above are the simplest functions that satisfy the constraints on these functions stated in the last sub-section. Next, we gather data from industry wide studies to obtain sample estimates for parameters.

As stated in Section 2, values of direct and indirect costs of violation (average of R_{int} and R_{ext} is \$300 in healthcare [46], a detailed breakdown is present in the ANSI report [2]), maximum personal benefit I (\$50 for medical records [2, 17]), etc. are available in studies. We assume $I_{max} = \$50$. Also, in absence of studies quantitatively distinguishing externally and internally caught violations we assume $R_{int} = R_{ext} = \$300$. Many parameters depends on the employee, his role in the organization and type of violation. Keeping a track of violations and behavior within the organization offers a data source for estimating and detecting changes in these parameters. We choose values for these parameters that are not extremes, $e = \$10$, $I = \$6$, $\epsilon_{th} = 0.03$, $\delta_{\mathcal{A}} = 0.4$ and $U_k = 40$. Further, under certain assumptions we calculate P_f (in Appendix B.2) to get $P_f = \$10$. Finally, the average cost of auditing C and

performance factor μ of log analysis tool should be known to $\mathcal{D}$. We assume values $C = \$50$, and an intermediate performance $\mu = 1.5$ of the tool.

Finally, analyzing data to detect and estimate may require the use of statistical methods, data mining and learning techniques. For example, consider detecting change in I. The expected behavior of $\mathcal{A}$ is determined by the equilibrium of the game (shown in next section), but, there is also the possibility of deviation with probability ϵ_{th}. Thus, this is a standard detection problem: a history of fixed finite length of employee actions can be used to obtain an estimate $\hat{\epsilon}_{th}$ and if the difference in $\hat{\epsilon}_{th}$ and ϵ_{th} is statistically significant then it can be claimed that I has changed. Various other methods [25] other than the simple one stated above can be used. We do not delve into details of these methods as that is beyond the scope of this paper and estimating risk parameters has been studied extensively in many contexts [2, 30, 40, 44, 49]. Later, in Section 5, we present a novel learning method to estimate I.

4 Equilibrium

In this section, we define a suitable equilibrium concept for the audit game (Section 4.1) and present an approximately cost-optimal strategy for the defender such that the best response to that strategy by the adversary results in the equilibrium being attained (Section 4.2). Recall that the equilibrium of the game occurs in the period in which the game parameters are fixed.

4.1 Equilibrium Concepts

We begin by introducing standard terminology from game theory. In a one-shot extensive form game players move in order. We assume player 1 moves first followed by player 2. An extensive form repeated game is one in which the round game is a one-shot extensive game. The history is a sequence of actions. Let H be the set of all possible histories. Let S_i be the action space of player i. A strategy of player i is a function $\sigma_i : H_i \to S_i$, where $H_i \subset H$ are the histories in which player i moves. The utility in each round is given by $r_i : S_1 \times S_2 \to \mathbb{R}$. The total utility is a δ_i-discounted sum of utilities of each round, normalized by $1 - \delta_i$.

The definition of strategies extends to extensive form repeated games with public signals. We consider a special case here that resembles our audit game. Player 1 moves first and the action is observed by player 2, then player 2 moves, but, that action may not be perfectly observed, instead resulting in a public signal. Let the space of public signals be Y. In any round, the observed public signal is distributed according to the distribution $\Delta Y(.|s)$, i.e., $\Delta Y(y|s)$ is the probability of seeing signal y when the action profile s is played. In these games, a history is defined as an alternating sequence of player 1'a action and public signals, ending in a public signal for histories in which player 1 moves and ending in player 1's move for histories in which player 2 moves. The actual utility in each round is given by the function $r_i : S_i \times Y \to \mathbb{R}$. The total expected utility g_i is the expected normalized δ_i-discounted sum of utilities of each round, where the expectation is taken over the distribution over public signals and histories. For any history h, the game to be played in the future after h is called the *continuation game* of h with total utility given by $g_i(\sigma, h)$.

A strategy profile (σ_1, σ_2) is a *subgame perfect equilibrium* (SPE) of a repeated game if it is a Nash equilibrium for all continuation games given by any history h [21]. One way of determining if a strategy is a SPE is to determine whether the strategy satisfies the *single stage deviation* property, that is, any *unilateral deviation* by any player *in any single round* is not profitable. We define a natural extension of SPE, which we call *asymmetric subgame perfect equilibrium* (or (ϵ_1, ϵ_2)-SPE), which encompasses SPE as a special case when $\epsilon_1 = \epsilon_2 = 0$.

Definition 1. *((ϵ_1, ϵ_2)-SPE) Denote concatenation operator for histories as ;. Strategy profile σ is a (ϵ_1, ϵ_2)-SPE if for history h in which player 1 has to play, given $h' = h; \sigma_1(h)$ and $h'' = h; s_1$,*

$$E(r_1(\sigma_1(h), \mathbf{y}))[\sigma_1(h), \sigma_2(h')] + \delta_1 E(g_1(\sigma, h'; \mathbf{y}))[\sigma_1(h), \sigma_2(h')] \geq E(r_1(s_1, \mathbf{y}))[s_1, \sigma_2(h'')] + \delta_1 E(g_1(\sigma, h''; \mathbf{y}))[s_1, \sigma_2(h'')] - \epsilon_1$$

for all s_1. For history h in which player 2 has to play, given $a(h)$ is the last action by player 1 in h, for all s_2

$$E(r_2(\sigma_2(h), \mathbf{y}))[a(h), \sigma_2(h)] + \delta_2 E(g_2(\sigma, h; \mathbf{y}))[a(h), \sigma_2(h)] \geq E(r_2(s_2, \mathbf{y}))[a(h), s_2] + \delta_2 E(g_2(\sigma, h; \mathbf{y}))[a(h), s_2] - \epsilon_2$$

We are particularly interested in $(\epsilon_1, 0)$-SPE, where player 1 is the defender and player 2 is the adversary. By setting $\epsilon_2 = 0$, we ensure that a rational adversary will never deviate from the expected equilibrium behavior. Such equilibria are important in security games, since $\epsilon_2 > 0$ could allow the adversary to deviate from her optimal strategy only for the purpose of causing significant loss to the defender. Also, in an $(\epsilon_1, 0)$-SPE, the defender is guaranteed to be at most within ϵ_1 of her optimal cost (which is the cost corresponding to $(0, 0)$-SPE), which is particularly relevant for $\mathcal{D}$ in the audit game, since $\mathcal{D}$ is rational and budget constrained.

The following useful property about history-independent strategies, which follows directly from the definition, helps in understanding our proposed history-independent audit strategy.

Property 1. *If a strategy profile σ is history-independent, i.e., $\sigma_1(h) = \sigma_1()$ and $\sigma_2(h) = \sigma_2(a(h))$ then the condition to test for SPE reduces to $E(r_1(\sigma_1(), \mathbf{y})) \geq E(r_1(s_1, \mathbf{y}))$, for player 1 and to $E(r_2(\sigma_2(h), \mathbf{y})) \geq E(r_2(s_2, \mathbf{y}))$, for player 2, since $g_i(\sigma, h; y)$ is the same for all y and each i. Also, if $E(r_i(s_i, \mathbf{y})) - E(r_i(\sigma_i(h), \mathbf{y})) \leq \epsilon_i$ for all i and s_i then σ is an (ϵ_1, ϵ_2)-SPE strategy profile.*

4.2 Equilibrium in the Audit Game

We next state an equilibrium strategy profile for the game $G_{\mathcal{A},k}$ such that $\mathcal{D}$ performs almost cost-optimally. Formally, we present a $(\epsilon_{\mathcal{A},k}, 0)$-SPE strategy profile, and calculate the value $\epsilon_{\mathcal{A},k}$. We also prove that the strategy profile is within $\sum_{\mathcal{A},k} \epsilon_{\mathcal{A},k}$ of the optimal cost of auditing for the organization. We accordingly refer to this strategy profile as a *near-optimal* strategy profile (for $\mathcal{D}$).

It is important that $\mathcal{D}$ makes its strategy publicly known, and provide a means to verify that it is playing that strategy. Indeed, the first mover in an extensive form game has an advantage in deciding the outcome of the game. Players can avail of this advantage by committing to their strategies, such as in inspection games [21]. As noted earlier, even though $\mathcal{D}$ acts after $\mathcal{A}$ does, yet, by committing to its strategy with a verification mechanism $\mathcal{D}$ simulates a first move by making the employee believe its commitment with probability 1. This also removes any variability in belief that different employees may have about the organization's strategy. Thus, we envision the organization making a commitment to stick to its strategy and providing a proof of the following the strategy. We argue that $\mathcal{D}$ will be willing to do so because (1) the equilibrium is $(\epsilon_1, 0)$-SPE so $\mathcal{A}$ is not likely to deviate, (2) $\mathcal{D}$ is patient and can bear a small loss due to occasional mistakes (with probability ϵ_{th}) by $\mathcal{A}$ and (3) the equilibrium we propose is close to optimal cost for $\mathcal{D}$, hence, the organization would be willing to commit to it and the employees would believe the commitment. Further, $\mathcal{D}$ making its strategy publicly known follows the general security principle of not making the security mechanisms private [51].

The main idea behind the definition of the near-optimal strategy profile is that $\mathcal{D}$ optimizes its utility assuming the best response of $\mathcal{A}$ for a given a^t. That is, $\mathcal{D}$ assumes that $\mathcal{A}$ does not commit any violations when (P, α) is in the deterred region, and systematically commits a violation otherwise (i.e., all of $\mathcal{A}$'s tasks are violations). Further, $\mathcal{D}$ assumes the worst case when the employee (with probability ϵ_{th}) accidentally makes a mistake in the execution of their strategy; in such a case, $\mathcal{D}$ expects all of $\mathcal{A}$'s tasks to be violations, regardless of the values of (P, α). This is because the distribution D_0^t over violations when $\mathcal{A}$ makes a mistake is unknown.

In other words, the expected cost function that $\mathcal{D}$ optimizes (for each total number of tasks a^t) is a linear sum of $(1 - \epsilon_{th})$ times the cost due to best response of $\mathcal{A}$ and ϵ_{th} times the cost when $\mathcal{A}$ commits all violations. The expected cost function is different in the deterred and non-deterred region due to the difference in best response of $\mathcal{A}$ in these two regions. The boundary between the deterred and non-deterred regions is conditioned by the value of the adversary's personal benefit I. We assume that $\mathcal{D}$ learns the value of the personal benefit within an error δI of its actual value, and that $\mathcal{D}$ does not choose actions (P, α) in the region of uncertainty determined by the error δI.

Formally, the expected reward is $E(\mathbf{Rew}_{\mathcal{D}}^t)[0]$ when the adversary commits no violation, and $E(\mathbf{Rew}_{\mathcal{D}}^t)[a^t]$ when all a^t tasks are violations. Both of these expected rewards are functions of P, α; we do not make that explicit for notational ease. Denote the deterred region determined by the parameter I and the budget as R_D^I and the non-deterred region as R_{ND}^I. Either of these regions may be empty. Denote the region (of uncertainty) between the curves determined by $I + \delta I$ and $I - \delta I$ as $R_{\delta I}^I$. Then the reduced deterred region is given by $R_D^I \backslash R_{\delta I}^I$ and the reduced non-deterred region by $R_{ND}^I \backslash R_{\delta I}^I$. The near-optimal strategy we propose is:

• For each possible number of tasks a^t that can be performed by $\mathcal{A}$, $\mathcal{D}$ using budget $b^t_{\mathcal{A},k}$, assumes the expected utility

$$U_D(P,\alpha) = (1-\epsilon_{th})E(\mathbf{Rew}^t_{\mathcal{D}})[0] + \epsilon_{th}E(\mathbf{Rew}^t_{\mathcal{D}})[a^t]\ ,$$

in $R^I_D \backslash R^I_{\delta I}$ and

$$U_{ND}(P,\alpha) = (1-\epsilon_{th})E(\mathbf{Rew}^t_{\mathcal{D}})[a^t] + \epsilon_{th}E(\mathbf{Rew}^t_{\mathcal{D}})[a^t]\ ,$$

in $R^I_{ND} \backslash R^I_{\delta I}$. $\mathcal{D}$ calculates the maximum expected utility across the two regions as follows:
– $U^D_{\max} = \max_{(P,\alpha)\in R^I_D \backslash R^I_{\delta I}} U_D(P,\alpha)$
– $U^{ND}_{\max} = \max_{(P,\alpha)\in R^I_{ND} \backslash R^I_{\delta I}} U_{ND}(P,\alpha)$
– $U = \max(U^D_{\max}, U^{ND}_{\max})$
$\mathcal{D}$ commits to the corresponding maximizer (P,α) for each a^t.
After knowing a^t, $\mathcal{D}$ plays the corresponding (P,α).

• $\mathcal{A}$ plays her best response (based on the committed action of $\mathcal{D}$), i.e., if she is deterred for all a^t she commits no violations and if she is not deterred for some a^t then all her tasks are violations, and she chooses the a^t that maximizes her utility from violations. But, she also commits mistakes with probability ϵ_{th}, and then the action is determined by distribution D^t_0.

Let $U^{D+\delta I}_{\max} = \max_{(P,\alpha)\in R^I_D \cup R^I_{\delta I}} U_D(P,\alpha)$, and $U^{ND+\delta I}_{\max} = \max_{(P,\alpha)\in R^I_{ND} \cup R^I_{\delta I}} U_{ND}(P,\alpha)$. We have the following result:

Theorem 1. *The near-optimal strategy profile (defined above) is an $(\epsilon_{\mathcal{A},k}, 0)$-SPE for the game $G_{\mathcal{A},k}$, where $\epsilon_{\mathcal{A},k}$ is*

$$\max\left(\max_{v^t,a^t}(U^{D+\delta I}_{\max} - U^D_{\max}),\ \max_{v^t,a^t}(U^{ND+\delta I}_{\max} - U^{ND}_{\max})\right) + \epsilon_{th}\max_{\alpha\in[0,1]}\left(\sum_{j=0}^{m-1}\delta^j_{\mathcal{D}}E(r(\vec{\mathbf{O}}^t, j))[U_k, U_k, \alpha]\right)$$

Remark 1. *The analysis of the theorem above is stable to errors in the value of the parameter, i.e., if the value of any parameter is wrong the analysis stays the same resulting in an $(\epsilon, 0)$-SPE, but, with ϵ greater than $\epsilon_{\mathcal{A},k}$.*

The proof is in Appendix B. The proof involves showing that the near-optimal strategy profile has the single stage deviation property. That $\mathcal{A}$ does not profit from deviating is immediate because $\mathcal{A}$ chooses the best response in each round of the game. The bound on profit from deviation for $\mathcal{D}$ has two terms. The first term arises due to $\mathcal{D}$ ignoring the region of uncertainty in maximizing its utility. The maximum difference in utility for the deterred region is $\max_{v^t,a^t}(U^{D+\delta I}_{\max} - U^D_{\max})$ and for the undeterred region is $\max_{v^t,a^t}(U^{ND+\delta I}_{\max} - U^{ND}_{\max})$. The first term is given by the maximum of these quantities. The second term arises due to the use of the worst case assumption of all (U_k) violations out of maximum possible U_k tasks when $\mathcal{A}$ makes a mistake as compared to the case when D^t_0 is known. Since $\mathcal{A}$'s choice only affects the violation loss part of $\mathcal{D}$'s utility and mistakes happen only with probability ϵ_{th}, the second term is the maximum possible loss due to violations multiplied by ϵ_{th}.

Numeric applications. The above theorem can be used to calculate concrete values for $\epsilon_{\mathcal{A},k}$ when all parametric functions are instantiated. For example, with the values in Section 3.3, we obtain $\epsilon_{\mathcal{A},k} = \$200$. Assuming $\mathcal{A}$ performs the maximum $U_k = 40$ number of tasks, $\epsilon_{\mathcal{A},k}$ is about 9.5% of the cost of auditing all actions of $\mathcal{A}$ with maximum punishment rate ($2100), with no violations, and about 3.3% of the cost incurred due to all violations caught externally ($6000), with no internal auditing or punishment. Similarly, if we assume 70% audit coverage with maximum punishment and four violations, the expected cost for organization is $2583, which means $\epsilon_{\mathcal{A},k}$ corresponds to about 7.7% of this cost. We present the derivation of value of $\epsilon_{\mathcal{A},k}$ in Claim 1 in Appendix B. The audit coverage here is for one employee only; hence it can be as high as 100%. Also, since $\mathcal{G}_{\mathcal{A}}$ is a parallel composition of the games $\mathcal{G}_{\mathcal{A},k}$ for all k, we claim that the near-optimal strategy profile followed for all games $\mathcal{G}_{\mathcal{A},k}$ is a $(\sum_k \epsilon_{\mathcal{A},k}, 0)$-SPE strategy profile for $G_{\mathcal{A}}$. (See Lemma 3 in Appendix B.1.) Finally, as noted earlier, the asymmetric equilibrium enables us to claim that the expected cost for $\mathcal{D}$ in the near-optimal strategy profile is at most $\sum_k \epsilon_{\mathcal{A},k}$ more than the optimal cost. Since costs add up linearly, $\mathcal{D}$'s cost in the whole audit process is at most $\sum_{\mathcal{A}}\sum_k \epsilon_{\mathcal{A},k}$ more than the optimal cost. Also, it follows from Remark 1 that errors in parameter estimates moves the cost of the audit process further away from the optimal cost.

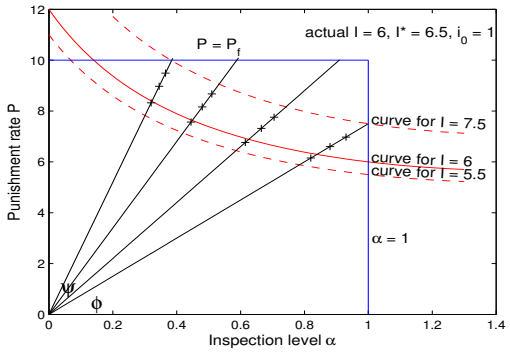

Figure 2: Visual representation of Algorithm 1, shown for $l = 4$ lines and $n = 3$ points($\times$) on the lines.

5 Learning Personal Benefit

In this section, we propose a learning mechanism to *learn* the personal benefit parameter I of $\mathcal{A}$, and prove that the algorithm is an effective learner. $\mathcal{D}$ has a prior belief about the value of I, namely $I \in [I^* - i_0, I^* + I_0]$ for parameters I^*, i_0. We assume the belief is correct. Call the intersection of the feasible region and the region between the curves given by $I^* - i_0$, $I^* + i_0$ the *search* region.

Algorithm 1 is the adversarial learning algorithm we propose to learn the separator given by $I = P^t(\nu(\alpha^t) + p(1 - \nu(\alpha^t)))$. The idea of the algorithm is to use a rotating sweep line [18] style technique, often used in geometric algorithmic problems. Multiple (l) lines passing through the origin and the search region are considered (see Figure 2). For any point P', α' there is a unique value I' such that $I' = P'(\nu(\alpha') + p(1 - \nu(\alpha')))$, then P', α' lies on the curve $I' = P(\nu(\alpha) + p(1 - \nu(\alpha)))$. Mark n points on each line within the search space, such that for any two consecutive point if I', I'' determines the curves these points lie on then $|I' - I''| = \frac{2i_0}{n+1}$. We call the two points nearest to the true separator 1-neighbors. The algorithm works by trying to find the non-deterred 1-neighbor point on each of the l lines.

We use the standard binary search algorithm, named BinSearch, as a black-box. In our setting BinSearch's queries are points on the line L_i. The point (P', α') lies in the detered region if and only if $I' \geq I$. The answer to the query $I' > I$ can therefore be inferred from the number of detected violations. However, there are two additional technical challenges that must be addressed (1) $\mathcal{A}$ may be willing to behave (not violate) in the non-deterred region (or misbehave in the deterred region) if doing so would bring rewards in the near future, and (2) even if $\mathcal{A}$ plays according to his immediate preferences during the learning phase the answers to BinSearch's queries are noisy due to the trembling hand assumption. We address the first challenge by ensuring that during any round of the learning phase the next d queries are not affected by $\mathcal{A}$'s current actions. This ensures that $\mathcal{A}$ always reveals her true preferences for points other than 1-neighbor points (Lemma 1). We address the second challenge (noisy binary search problem) by querying points multiple times and taking the majority vote to ensure that the results are accurate with high probability(Lemma 2).

We make multiple copies of BinSearch, such that BinSearch_i searches on line L_i. BinSearch returns the desired point in the field $Result$, which can be null in case there is no such point. A field $Done$ checks if BinSearch is done with its processing and $allDone$ checks whether all copies of BinSearch are done with processing. If some copy of BinSearch finishes before others a dummy point (α'_0, P'_0) may be queried to ensure that the next d queries remain fixed.. Each query point is used $2q+1$ times for auditing before using a majority vote to determine the answer. Finally, d dummy rounds of audit are performed with points $(\alpha'_1, P'_1)...(\alpha'_d, P'_d)$ chosen before the algorithm starts. The first step in analyzing Algorithm 1 is proving that the optimal behavior of $\mathcal{A}$ is playing the best response for most rounds, which is not true in general for repeated games.

Lemma 1. *Assume $(2q + 1)(l - 1) \geq d$. Then in Algorithm 1 the adversary chooses the best response for all points on each line, except the 1-neighbors, if $d \geq \log_{1/\delta_{\mathcal{A}}} \frac{(n+1)(P_f U_k + I_{max} U_k)}{(1-\delta_{\mathcal{A}}^2)2i_0}$.*

The key idea in the proof is that in any non-dummy round the next d queries are fixed and known to $\mathcal{A}$, which is ensured by the d dummy rounds at the end. So only after at least d rounds does the adversary earn any benefit from

Algorithm 1 Adversarial learning

Require: $n, l, d, q, I^*, i_o, (\alpha'_0, P'_0), ..., (\alpha'_d, P'_d), (2q+1)(l-1) \geq d$.
  $\theta \leftarrow \frac{\psi}{(l-1)}$, L_i is the line through $(0,0)$ at angle $\phi + (i-1)\theta$.
  Let $d_i = 2i_0/(n+1)$
  for $i = 1$ **to** l **do**
    Mark n points on L_i in the search space, so that the j^{th} $(1 \leq j \leq n)$ point lies on the curve given by $I^* - i_0 + jd_i$.
  end for
  Make l copies of BinSearch: $\mathsf{BinSearch}_1, ..., \mathsf{BinSearch}_l$
  while $allDone \neq true$ **do**
    for $i = 1$ **to** l **do**
      $deter \leftarrow 0$, $(\alpha_i^x, P_i^x) \leftarrow (\alpha'_0, P'_0)$ {default query}
      if $\mathsf{BinSearch}_i.Done \neq true$ **then**
        $(\alpha_i^x, P_i^x) \leftarrow \mathsf{BinSearch}_i.nextQuery$
      end if
      for $k = 1$ **to** $2q+1$ **do**
        Play the point (α_i^x, P_i^x) in the audit game.
        if number of detected violations $= 0$ **then**
          $deter \leftarrow deter + 1$
        end if
      end for
      if $deter \geq q+1$ **then**
        Return deterred to $\mathsf{BinSearch}_i$
      else
        Return non-deterred to $\mathsf{BinSearch}_i$
      end if
      if $\mathsf{BinSearch}_i.Done = true$ **then**
        $(\alpha_i, P_i) \leftarrow \mathsf{BinSearch}_i.Result$
      end if
    end for
    $allDone \leftarrow \mathsf{BinSearch}_1.Done \wedge .. \wedge \mathsf{BinSearch}_l.Done$
  end while
  Play the point $(\alpha'_1, P'_1), ..., (\alpha'_d, P'_d)$ in the audit game.
  Use the curve (of the form of separator) that passes through majority of the points $(\alpha_1, P_1), ..., (\alpha_l, P_l)$ to learn I.

not playing the best response. Since that the employee is not patient (small $\delta_{\mathcal{A}}$), thus, the utility earned in future is not very important. By making d large enough the benefit earned in future by not playing best response is not significant for $\mathcal{A}$. The proof is in Appendix B. Next, we claim that Algorithm 1 is an effective learner.

Theorem 2. *Assume that all conditions and result of Lemma 1 hold. In Algorithm 1, the defender learns the value of I with error bounded by $4i_0/(n+1)$ and probability greater than*

$$\left(1 - \sum_{i=q+1}^{2q+1} \binom{2q+1}{i} (\epsilon_{th})^i (1-\epsilon_{th})^{2q+1-i}\right)^{l\lceil \log_2 n \rceil}$$

in at-most $l(2q+1)\lceil \log_2 n \rceil + d$ number of rounds.

The proof is in Appendix B.The proof involves using the majority vote technique for the l lines with the number of BinSearch's queries no larger than $\lceil \log_2 n \rceil$ to get the high probability bound. Then, since any two consecutive points on each line lie on curves given by I', I'' such that $|I' - I''| = \frac{2i_0}{n+1}$, we obtain the error bound in the learned parameter.

Observe that choosing a large n will ensure a small error in the learned value of I. Also, using Hoeffding inequality [27] the probability bound above is $\geq (1 - \exp(-2(2q+1)(0.5 - \epsilon_{th})^2))^{l(\lfloor \log_2 n \rfloor + 1)}$, which is higher for higher values of q. But, a large n and q also increases the running time of the algorithm, which in practice should not be large. Thus, operationally there is a balance required in the choice of n and q. We show some concrete values for Theorem 2. Using $i_0 = 1, n = 7$, in addition to values from Section 3.3, in Lemma 1 we get $d \geq 10.19$, thus, choose $d = 11$. Choose $l = 5, q = 1$ to satisfy $(2q+1)(l-1) \geq d$. Then, Algorithm 1 produces a value for I with the error bound 0.5 and probability 0.96 in 56 rounds. Thus, with daily audits the adversary's personal benefit is learned in 2 months with near certainty.

Before concluding the section, we discuss the algorithm above in the context of two areas in algorithm design: mechanism design and noisy binary search. First, mechanism design is a technique to design game strategy and incentives so that players with a private type (I in our case) reveal the true value of their type. Its use varies widely from auctions to selling goods. In online mechanism design with fixed types, commitment to a strategy by the mechanism designer is required for truthful revelation of type [43]. If $\mathcal{D}$ commits to playing all actions in the learning phase then our algorithm can be considered to be an instance of approximate [52] online mechanism design [43] with player $\mathcal{A}$ and mechanism designer $\mathcal{D}$, i.e., $\mathcal{A}$ has incentives to reveal his type I with δI deviation. The approximation arises because $\mathcal{D}$ cannot learn I perfectly from the finite learning phase. However, practically $\mathcal{D}$'s commitment would not be credible, as $\mathcal{D}$ would want to deviate to the more cost-optimal equilibrium audit strategy than keep playing the learning phase after knowing I.

We discuss other algorithms for noisy binary search that can be plugged into Algorithm 1 to learn the adversary's personal benefit. These algorithms take longer than the binary search in Algorithm 1 to learn while guaranteeing that the learned value is correct with higher probability. We favor the simple binary search we use since 1) for the choice of parameters above it has low running time $3 \log n$ and 2) the probability bound we obtain is acceptable for learning with humans. The binary search by Nowak [41] is based on multiplicative weight update and takes time more than $\frac{4 \log(n/\delta)}{1 - \sqrt{2\epsilon(1-\epsilon)}}$ for probability of success $1 - \delta$ with error rate ϵ. The algorithm by Karp et al. [31] resembles standard binary search, except that it allows backtracking: the algorithm checks if the current sub-interval is the right one else it backtracks to the larger parent interval, using at least 3 queries in the interval. However, the check has non-zero probability of error even on the right interval, thus, the expected running time of this algorithm is more than $3 \log n$, but, the probability of success is higher than the simple binary search. Thus, since running time is important, the simple binary search we use is best for our case. However, if the probability of success is critical then Algorithm 1 can use any binary search algorithm as long as next d steps can be determined by the adversary, so that she plays the best response (Lemma 1).

As a final comment, the strength of the algorithm above lies in the fact that it can learn any separator (with small error) by choosing l properly. Thus, even if we do not assume any knowledge of the separator we can still use the algorithm above to learn the regions.

6 Predictions and Interventions

In this section, we use our model to predict observed practices in industry and the effectiveness of policy interventions in encouraging organizations to conduct more thorough audits by analyzing the equilibrium audit strategy P, α under varying parameters. We use the values of parameters and instantiation of functions given in Section 3.3 (unless otherwise noted). We assume that the value of personal benefit I is learned exactly and that P and α take discrete values, with the discrete increments being 0.5 and 0.05, respectively. We also assume for sake of exposition that $u_k = U_k$, i.e., the number of tasks is fixed.

Average cost R_{ext} and probability p of external detection of violation. We vary R_{ext} from \$5 to \$3900, with R_{int} fixed at \$300. The results are shown in Figure 3. There are two cases shown in the figure: $p = 0.5$ and $p = 0.9$. The figure shows the equilibria P, α chosen for different values of R_{ext}.

Prediction 1: Increasing R_{ext} and p is an effective way to encourage organizations to audit more. In fact, when $p * R_{ext}$ is low X may not audit at all. Thus, X audits to protect itself from greater loss incurred when violations are caught externally. Surprisingly, the hospital may continue to increase inspection levels (incurring higher cost) beyond the minimum level necessary to deter a rational employee. Hospital X does so because the employee is not fully

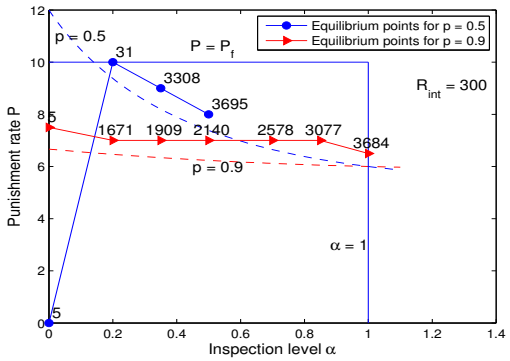

Figure 3: The dashed separator for 2 values of p, and equilibrium (see legend) P, α for varying values (shown above point) of R_{ext} from 5 to 3900.

rational: even in the deterred region there is an ϵ_{th} probability of violations occurring.

Suggested Intervention 1: Subject organizations to external audits and fines when violations are detected. For example, by awarding contracts for conducting 150 external audits by 2012 [26], HHS is moving in the right direction by effectively increasing p. This intervention is having an impact: the 2011 Ponemon study on patient privacy [47] states—"Concerns about the threat of upcoming HHS HIPAA audits and investigation has affected changes in patient data privacy and security programs, according to 55 percent of respondents."

Prediction 2: Interventions that increase the expected loss for both external and internal detection of violations are not as effective in increasing auditing as those that increase expected loss for external detection of violations only. Table 5 shows the equilibrium inspection level as R_{ext} and R_{int} are both increased at the same rate. While the inspection level may initially increase, it quickly reaches a peak. As an example, consider the principle of breach detection notification used in many data breach laws [50]. The effect of breach detection notification is to increase both R_{int} and R_{ext} since notification happens for all breaches. While there isn't sufficient data for our model to predict whether these laws are less effective than external audits (see suggested study below), prior empirical analysis [50] indicate that the benefit in breach detection from these laws is only about 6% (after adjusting for increased reporting of breaches due to the law itself).

Suggested study: An empirical study that separately reports costs incurred when violations are internally detected from those that are externally detected would be useful in quantifying and comparing the effectiveness of interventions. Existing studies either do not speak of these distinct categories of costs [46, 50] or hint at the importance of this distinction without reporting numbers [45, 57].

Punishment loss factor e and personal benefit I. *Prediction 3: Employees with higher value for e (e.g., doctors have higher e; suspending a doctor is costlier for the hospital than suspending a nurse) will have lower punishment levels.* If punishments were free, i.e., $e = 0$, (an unrealistic assumption) X will always keep the punishment rate at maximum according to our model. At higher punishment rates ($e = 1000$), X will favor increasing inspections rather than increasing the punishment level P (see Table 1 in Appendix A). While we do not know of an industry-wide study on this topic, there is evidence of such phenomena occurring in hospitals. For example, in 2011 Vermont's Office of Professional Regulation, which licenses nurses, investigated 53 allegations of drug diversion by nurses and disciplined 20. In the same year, the Vermont Board of Medical Practice, which regulates doctors, publicly listed 11 board actions against licensed physicians for a variety of alleged offenses. However, only one doctor had his license revoked while the rest were allowed to continue practicing [33].

Prediction 4: Employees who cannot be deterred are not punished. When the personal benefit of the employee I is high, our model predicts that X chooses the punishment rate $P = 0$ (because this employee cannot be deterred at all) and increases inspection as R_{ext} increases to minimize the impact of violations by catching them inside (see Table 2 in Appendix A). Note that this is true only for violations that are not very costly (as is the case for our choice of costs). If the expected violation cost is more than the value generated by the employee, then it is better to fire the non-deterred employee (see Appendix B.2).

When I is low, the employee is deterred even for low values of P, α. While this seems good for X, an important consideration is the scenario of *trust trap* [5]. In a trust trap the employee earns the trust of X by behaving properly for many audit cycles, and then commits a costly violation and leaves. Our model predicts trust trap.

Prediction 5: If only the history of employee actions is used to learn I and there have been no past violations then the value of learned I will be small. A small I will mean that X will select lower inspection levels. This would enable a patient (and devious!) employee to get away with costly violations.

Suggested Intervention 5: X can make informed decisions to avoid the trust trap, e.g., set a minimum value for I.

Audit cost C and performance factor μ of log analysis tool.

Prediction 6: If audit cost C decreases or the performance μ of log analysis increases, then the equilibrium inspection level increases. The data supporting this prediction is presented in Table 3 and 4 in Appendix A. Intuitively, it is expected that if the cost of auditing goes down then organizations would audit more, given their fixed budget allocated for auditing. Similarly, a more efficient mechanized audit tool will enable the organization to increase its audit efficiency with the fixed budget. For example, MedAssets claims that Stanford Hospitals and Clinics saved about $\$4$ million by using automated tools for auditing [38].

7 Related Work

Auditing and Accountability: Prior work studies orthogonal questions of algorithmic detection of policy violations [6, 8, 22, 54] and blame assignment [4, 7, 29, 34]. Feigenbaum et al. [20] report work in progress on formal definitions of accountability capturing the idea that violators are punished with or without identification and mediation with non-zero probability, and punishments are determined based on an understanding of "typical" utility functions. Operational considerations of how to design an accountability mechanism that effectively manages organizational risk is not central to their work. In other work, auditing is employed to revise access control policies when unintended accesses are detected [9, 35, 56]. Another line of work uses logical methods for enforcing a class of policies, which cannot be enforced using preventive access control mechanisms, based on evidence recorded in audit logs [13]. Cheng et al. [15,16] extend access control to by allowing agents access based on risk estimations. A game-theoretic approach of coupling access control with audits of escalated access requests in the framework of a single-shot game is studied by Zhao et al. [59]. These works are fundamentally different from our approach. We are interested in scenarios where access control is not desirable and audits are used to detect violations. We believe that a repeated game can better model the repeated interactions of auditing.

Risk Management and Data Breaches: Our work is an instance of a risk management technique [40, 44] in the context of auditing and accountability. As far as we know, our technique is the first instance of managing risk in auditing using a repeated game formalism. Risk assessment has been extensively used in many areas [30, 49]; the report by American National Standards Institute [2] provides a risk assessment mechanism for healthcare. Our model also models data breaches that happen due to insider attacks. Reputation has been used to study insider attacks in non-cooperative repeated games [58]; we differ from that work in that the employer-employee interaction is essentially cooperative. Also, the primary purpose of interaction between employer and employee is to accomplish some task (e.g., provide medical care). Privacy is typically a secondary concern. Our model captures this reality by considering the effect of non-audit interactions in parameters like P_f. There are quite a few empirical studies on data breaches and insider attacks [46,50,57] and qualitative models of insider attacks [5]. We use these studies to estimate the parameters in our model and to evaluate the predictions of our model.

Adversarial Learning: One of the earliest works on adversarial machine learning is by Kearns et al. [32] on extending the PAC learning model to allow a fixed probability of labeling error. Auer et al. [3] extend the online model to allow for bounded malicious error. In these works, the adversary can always fool the learner unless there is a constraint on the number of labels she can label wrongly. In contrast, in our setting training data points are *chosen* by the learner (organization) and the *labels provided* by the adversary (employee) in an online manner. The learner outputs the separator after all the training data has been collected. We use a novel reasoning involving delayed benefits for the impatient adversary in a repeated game setting to impose bounds on the malicious error.

Lowd and Meek [36] study the problem of learning in an adversarial setting by proposing a framework to study reverse engineering of classifiers to perform cost-optimal (cost of adversary) evasion in reasonable amount of time.

They further propose a classifier modification that predicts the adversary's evasion based on the assumption of myopic adversary actions and adapts the classifier to counter the evasion. In particular, they do not look for an equilibrium of this repeated game. Nelson et al. [39] provide models of attacks on machine learning algorithms, and demonstrate a few attacks. They also improve upon algorithms to find cost-optimal evasions by the attacker. In contrast, our approach uses the adversary's discounting of future benefit to allow the learner to incentivize the adversary to behave in a desired manner. Nelson et al. [39] also propose using *regret minimizing* [11] technique in case of repeated game learning setting, which converges to the best classifier in hindsight w.r.t. to a given set of classifiers with no assumption about the adversary. In contrast, we model the almost rational adversary's unknown incentives and provide high probability guarantees of learning.

8 Conclusion

First, as public policy and industry move towards accountability-based privacy governance, the biggest challenge is how to operationalize requirements such as internal enforcement of policies. Principled audit and punishment schemes like the one presented in this paper will be part of the enforcement regime making these results significant in practice. Second, a usual complaint against this kind of risk management approach is that there isn't data to estimate the risk parameters. We provide evidence that a number of parameters in the game model can be estimated from prior empirical studies, suggest specific studies that can help estimate other parameters, and design a learning algorithm that the defender can use to provably learn the adversary's private incentives. Moving forward, we plan to generalize our results to account for colluding adversaries, explore the space of effective policy interventions, and evaluate these mechanisms through user studies.

References

[1] *HIPAA Enforcement*, 2012 (accessed May 1,2012).

[2] American National Standards Institute (ANSI) / The Santa Fe Group /Internet Security Alliance. The financial impact of breached protected health information, 2012 (accessed May 1,2012).

[3] P. Auer and N. Cesa-Bianchi. On-line learning with malicious noise and the closure algorithm. *Annals of Mathematics and Artificial Intelligence*, 23:83–99, 1998.

[4] M. Backes, A. Datta, A. Derek, J. C. Mitchell, and M. Turuani. Compositional analysis of contract-signing protocols. *Theor. Comput. Sci.*, 367(1-2):33–56, 2006.

[5] S. R. Band, D. M. Cappelli, L. F. Fischer, A. P. Moore, E. D. Shaw, and R. F. Trzeciak. Comparing Insider IT Sabotage and Espionage: A Model-Based Analysis. Technical Report CMU/SEI-2006-TR-026, CMU, December 2006.

[6] A. Barth, A. Datta, J. C. Mitchell, and H. Nissenbaum. Privacy and contextual integrity: Framework and applications. In *IEEE Symposium on Security and Privacy*, pages 184–198, 2006.

[7] A. Barth, A. Datta, J. C. Mitchell, and S. Sundaram. Privacy and utility in business processes. In *CSF*, pages 279–294, 2007.

[8] D. A. Basin, F. Klaedtke, and S. Müller. Policy monitoring in first-order temporal logic. In *CAV*, pages 1–18, 2010.

[9] L. Bauer, S. Garriss, and M. K. Reiter. Detecting and resolving policy misconfigurations in access-control systems. In *SACMAT*, pages 185–194, 2008.

[10] J. Blocki, N. Christin, A. Datta, and A. Sinha. Regret minimizing audits: A learning-theoretic basis for privacy protection. In *Proceedings of the 24th IEEE Computer Security Foundations Symposium, CSF 2011*, pages 312–327, 2011.

[11] A. Blum and Y. Mansour. *Algorithmic Game Theory*, chapter Learning, regret minimization, and equilibria, pages 79–102. Cambridge University Press, 2007.

[12] Casey Ichniowski and Kathryn Shaw and Giovanna Prennushi. The Effects of Human Resource Management Practices on Productivity. Technical Report 5333, National Bureau of Economic Research, November 1995.

[13] J. G. Cederquist, R. Corin, M. A. C. Dekker, S. Etalle, J. I. den Hartog, and G. Lenzini. Audit-based compliance control. *Int. J. Inf. Sec.*, 6(2-3):133–151, 2007.

[14] Center for Information Policy Leadership. Accountability-Based Privacy Governance Project, 2012 (accessed May 1,2012).

[15] P.-C. Cheng and P. Rohatgi. IT Security as Risk Management: A Research Perspective. *IBM Research Report*, RC24529, April 2008.

[16] P.-C. Cheng, P. Rohatgi, C. Keser, P. A. Karger, G. M. Wagner, and A. S. Reninger. Fuzzy Multi-Level Security : An Experiment on Quantified Risk-Adaptive Access Control. In *Proceedings of the IEEE Symposium on Security and Privacy*, 2007.

[17] Cole Petrochko. DHC: EHR Data Target for Identity Thieves, December 2011.

[18] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson. *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd edition, 2001.

[19] Fairwarning. Industry Best Practices for Patient Privacy in Electronic Health Records, April 2011.

[20] J. Feigenbaum, A. D. Jaggard, and R. N. Wright. Towards a formal model of accountability. In *Proceedings of the 2011 workshop on New security paradigms workshop*, 2011.

[21] D. Fudenberg and J. Tirole. *Game Theory*. The MIT Press, 1991.

[22] D. Garg, L. Jia, and A. Datta. Policy auditing over incomplete logs: theory, implementation and applications. In *Proceedings of the 18th ACM Conference on Computer and Communications Security, CCS 2011*, pages 151–162, 2011.

[23] F. L. Greitzer and R. E. Hohimer. Modeling human behavior to anticipate insider attacks. *Journal of Strategic Security*, IV:25–48, 2011.

[24] J. Grossklags, N. Christin, and J. Chuang. Secure or insure? A game-theoretic analysis of information security games. In *Proceedings of the 2008 World Wide Web Conference (WWW'08)*, pages 209–218, Beijing, China, Apr. 2008.

[25] E. A. Hanushek. *Statistical Methods for Social Scientists*. New York: Academic Press, 1977.

[26] HHS. HIPAA Privacy and Security Audit Program.

[27] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[28] G. Hulme. Steady Bleed: State of HealthCare Data Breaches, September 2010. InformationWeek.

[29] R. Jagadeesan, A. Jeffrey, C. Pitcher, and J. Riely. Towards a theory of accountability and audit. In *ESORICS*, pages 152–167, 2009.

[30] Karim H. Vellani. Strategic Healthcare Security, Risk Assessments in the Environment of Care, 2008. Report for Wisconsin Healthcare Engineering Association.

[31] R. M. Karp and R. Kleinberg. Noisy binary search and its applications. In *SODA*, pages 881–890, 2007.

[32] M. J. Kearns and M. Li. Learning in the presence of malicious errors. *SIAM Journal on Computing*, 22(4):807–837, 1993.

[33] Ken Picard. Are Drug-Stealing Nurses Punished More Than Doctors?, 2012.

[34] R. Küsters, T. Truderung, and A. Vogt. Accountability: definition and relationship to verifiability. In *ACM Conference on Computer and Communications Security*, pages 526–535, 2010.

[35] B. W. Lampson. Computer security in the real world. *IEEE Computer*, 37(6):37–46, 2004.

[36] D. Lowd and C. Meek. Adversarial learning. In *KDD*, pages 641–647, 2005.

[37] G. J. Mailath and L. Samuelson. *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press, USA, 2006.

[38] MedAssets. MedAssets Case Sudy: Stanford hospital takes charge of its charge capture process, increasing net revenue by 4 million, 2011.

[39] B. Nelson. *Behavior of Machine Learning Algorithms in Adversarial Environments*. PhD thesis, University of California, Berkeley, Nov. 2010.

[40] NIST. Guide for Conducting Risk Assessments, September 2011.

[41] R. D. Nowak. Noisy generalized binary search. In *NIPS*, pages 1366–1374, 2009.

[42] C. Ornstein. Breaches in privacy cost Kaiser, May 2009. Available online at: http://articles.latimes.com/2009/may/15/local/me-privacy15.

[43] D. C. Parkes. *Algorithmic Game Theory*, chapter Online Mechanisms, pages 411–439. Cambridge University Press, 2007.

[44] Pau-Chen Cheng and Pankaj Rohatgi. IT Security as Risk Management: A Reserach Perspective, April 2008. IBM Research Report.

[45] Ponemon Institute, LLC. Benchmark Study on Patient Privacy and Data Security, November 2010.

[46] Ponemon Institute, LLC. 2010 Annual Study: U.S. Cost of a Data Breach, March 2011.

[47] Ponemon Institute, LLC. Second Annual Benchmark Study on Patient Privacy and Data Security, December 2011.

[48] Ponemon Institute, LLC. 2011 Cost of Data Breach Study: United States, March 2012.

[49] PricewaterhouseCoopers. A practical guide to risk assessment, December 2008.

[50] S. Romanosky, D. Hoffman, and A. Acquisti. Empirical analysis of data breach litigation. In *ICIS*, 2011.

[51] J. H. Saltzer and M. D. Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975.

[52] J. Schummer. Almost-dominant strategy implementation. *GAMES AND ECONOMIC BEHAVIOR*, pages 154–170, 1999.

[53] The White House. Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy, 2012 (accessed May 1,2012).

[54] M. C. Tschantz, A. Datta, and J. M. Wing. Formalizing and enforcing purpose requirements in privacy policies. In *IEEE Symposium on Security and Privacy*, 2012. To Appear.

[55] H. Varian. System reliability and free riding. In L. Camp and S. Lewis, editors, *Economics of Information Security (Advances in Information Security, Volume 12)*, pages 1–15. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[56] J. A. Vaughan, L. Jia, K. Mazurak, and S. Zdancewic. Evidence-based audit. In *CSF*, pages 177–191, 2008.

[57] Verizon. 2012 Data Breach Investigations Report, 2012.

[58] N. Zhang, W. Yu, X. Fu, and S. K. Das. Towards effective defense against insider attacks: The establishment of defender's reputation. In *Proceedings of the 2008 14th IEEE International Conference on Parallel and Distributed Systems*, pages 501–508, 2008.

[59] X. Zhao and M. E. Johnson. Access governance: Flexibility with escalation and audit. In *HICSS*, pages 1–13, 2010.

A Data

R_{ext}	P	α
5 to 443	0	0
443 to 3900	6.5	1

Table 1: P, α for $e = 1000$

R_{ext}	P	α
5	0	0
670	0	0.1
685	0	0.35
714	0	0.6
748	0	0.85
790	0	1.0

Table 2: P, α for $I = 50$

C	P	α
10	6.5	1
20	6.5	1
30	7.0	0.85
40	7.5	0.65
50	8.0	0.5
60	9.5	0.25
70	10.0	0.2

Table 3: P, α for varying C

μ	P	α
1.0	10.0	0.3
1.2	9.5	0.35
1.3	9.5	0.35
1.40	9.0	0.45
1.5	9.0	0.45
1.6	8.5	0.5
1.7	8.5	0.5

Table 4: P, α for varying μ

R_{ext} and R_{int}	P	α
5 to 26	0	0
26 to 3900	10	0.2

Table 5: P, α for constant (0) difference in R_{int}, R_{ext}

B Proofs

Reminder of Lemma 1. *Assume $(2q+1)(l-1) \geq d$. Then in Algorithm 1 the near-rational adversary chooses the best response for all points on each line, except the 1-neighbors, if*

$$d \geq \log_{1/\delta_{\mathcal{A}}} \frac{(n+1)(P_f U_k + I_{max} U_k)}{(1-\delta_{\mathcal{A}}^2) 2 i_0} .$$

Proof. First, observe that an any round t the organizations action is known and fixed for the next $(2q+1)(l-1)$ rounds. This is because for the next $(2q+1)(l-1)$ the organization queries points on each line ($2q+1$ queries on each line) as would be asked by BinSearch. But, these queries of BinSearch are already known, as these are the next query for each lines which is determined by the history of queries for that line which is known exactly at time t.

Consider two possible choices of punishment and level of inspection P', α' and P, α in some future round. The absolute difference in expected utility in that round for number of violations v'^t and v^t in the two scenarios is $I|v^t - v'^t| + \max(v^t, v'^t)|P(\nu(\alpha) + p(1-\nu(\alpha))) - P'(\nu(\alpha') + p(1-\nu(\alpha')))|$. Since $P_f \geq P(\nu(\alpha) + p(1-\nu(\alpha))) \geq 0$ for any feasible P, $I \leq I_{max}$ and $\max(v^t, v'^t) \leq U_k$ the absolute difference in expected utility per round is bounded by $P_f U_k + I_{max} U_k$. Thus, $P_f U_k + I_{max} U_k$ is the maximum expected benefit that the adversary can get, by making the organization play differently.

Next, we show that not playing the best response for any non-1-neighbor points in the learning results in a total expected utility that is less than the total expected utility when a best response is played, thus, by rationality a rational adversary will always play the best response for the non-1-neighbor points in the learning, i.e., provide the right answer. Observe that not playing the best response means not committing all violations when $I - P(\nu(\alpha) + p(1-\nu(\alpha))) > 0$ or not committing 0 violations when $I - P(\nu(\alpha) + p(1-\nu(\alpha))) < 0$. Also, the minimum loss for not playing best response is when the deviation from number of violations is only one. Let P, α be any non-neighbor point. Then a difference of one violations for this point produces a difference in utility given by

$$L_{min} = |I - P(\nu(\alpha) + p(1-\nu(\alpha)))|$$

L_{min} is the least difference in expected utility in a round in the two scenarios: when the adversary does not play the best response and when he does play the best response, since his minimum deviation is by one violation. Now, the non-1-neighbor point is on the curve defined by value that differs from I by a minimum of $2i_0/(n+1)$. (by construction of algorithm). Thus, $L_{min} \geq 2i_0/(n+1)$. Also note that the majority vote for each point forces the adversary to provide at least two non-best responses to make any difference on the learning.

Remember that the learning algorithm is known to the adversary and the organization is committed to stick to this algorithm. Consider two scenarios: one in which the adversary decides to not play the best response (call it bad world) for the first time at round t_0 and one in which he always plays the best response (call it good world), for any point. Thus, before t_0 adversary's response is the best response in both worlds. In both worlds the organizations actions are known and fixed till at-least round $t_0 + d$. Then the best case for the adversary in the bad world is when only two non-best response (fooling the majority vote) gives him the expected benefit of $P_f U_k + I_{max} U_k$ after the $t_0 + d$ round. Thus, from time $t_0 + 1$ to at-least $t_0 + d$ the adversary's response is the best response in both worlds in such a case, since the learner is committed to its actions in these rounds. Since playing the best response is a history-independent strategy the difference in expected payoff from $t_0 + 1$ to $t_0 + d$ is 0 (using Property 1). We have already calculated the upper bound on the benefit in expected utility every round after at-least the round $t_0 + d$ as $P_f U_k + I_{max} U_k$. Thus, the difference in total expected utility in these two worlds (starting from round t_0) is at most

$$(1-\delta_{\mathcal{A}})(-L_{min} - \delta_{\mathcal{A}} L_{min}) + (\delta_{\mathcal{A}})^d (P_f U_k + I_{max} U_k).$$

Given $d \geq \log_{1/\delta_{\mathcal{A}}} \frac{(n+1)(P_f U_k + I_{max} U_k)}{(1-\delta_{\mathcal{A}}^2) 2 i_0}$ we have $(\delta_{\mathcal{A}})^d (P_f U_k + I_{max} U_k) \leq (1-\delta_{\mathcal{A}}^2) 2 i_0/(n+1)$. Thus, by rationality the adversary chooses the action with more utility at round t_0, and by definition that is the best response at round t_0. □

Lemma 2. *Assume all condition and result of Lemma 1 hold and that the adversary is near-rational. Let the number of queries asked by* BinSearch *for some given line among the l lines be N. The probability that the answers for non-1-neighbor points on the line are correct is greater than*

$$\left(1 - \sum_{i=q+1}^{2q+1} \binom{2q+1}{i} (\epsilon_{th})^i (1-\epsilon_{th})^{2q+1-i}\right)^N$$

Proof. First, we calculate the probability of a wrong answer for any non-1-neighbor point used in the audit game. Using Lemma 1 we already know that the adversary will provide his best response; i.e., she will choose 0 violations in deterred region and all violations in non-deterred region, but, can make a mistake with probability ϵ_{th} due to the trembling hand assumption. Thus, with probability $1-\epsilon_{th}$ the region detected will be the right region. We consider the worst case when the adversary makes a mistake. The worst case for a point in the deterred region is when the adversary commits all violations and for a point in the non-deterred region is when the adversary commits 0 violations; in both cases the probability of a wrong region being detected is 1. Thus, by the law of total probability, the probability of wrong region answer in one round is upper bounded by ϵ_{th}. Also, using majority vote with $2q+1$ points yields a wrong answer when more than q answers are wrong. Since all answers are independent, i answers are wrong with probability less than $\binom{2q+1}{i}(\epsilon_{th})^i(1-\epsilon_{th})^{2q+1-i}$, and thus, the probability of the majority vote being right is given by

$$\left(1 - \sum_{i=q+1}^{2q+1} \binom{2q+1}{i} (\epsilon_{th})^i (1-\epsilon_{th})^{2q+1-i}\right)$$

Then, because the same but independent procedure is repeated for the N points in the worst case (if the 1-neighbor points are queried k times then we have only $N-k$ queries) we obtain the desired result. □

Reminder of Theorem 2. *Assume that all conditions and result of Lemma 1 hold. In Algorithm 1, the defender learns the value of I with error bounded by $4i_0/(n+1)$ and probability greater than*

$$\left(1 - \sum_{i=q+1}^{2q+1} \binom{2q+1}{i} (\epsilon_{th})^i (1-\epsilon_{th})^{2q+1-i}\right)^{l\lceil \log_2 n \rceil}$$

in at-most $l(2q+1)\lceil \log_2 n \rceil + d$ number of rounds.

Proof. First, for every line the number of queries of BinSearch is not more than $\lceil \log_2 n \rceil$. Using Lemma 2 we can claim that with probability greater than

$$\left(1 - \sum_{i=q+1}^{2q+1} \binom{2q+1}{i} (\epsilon_{th})^i (1-\epsilon_{th})^{2q+1-i}\right)^{l\lceil \log_2 n \rceil}$$

all the non-1-neighbor answers for all l lines will be correct. In such a case, the algorithm cannot find any of the deterred 2 or higher neighbor points as the desired non-deterred point. Also it cannot find any of the non-deterred 3 or higher neighbor points as the desired point. Thus, it only finds the non-deterred 1 or 2 neighbor or deterred 1-neighbor point as the final answer. Then, the curve fitted in the final step of the algorithm will yield the value of I that corresponds to the curve on majority of the final answers of each line lie. By the restriction on the final answers this curve will either pass through non-deterred 1 or 2 neighbor or deterred 1-neighbor points for all lines. Also, the curve for the true value of I lies between non-deterred 1-neighbor and deterred 1-neighbor points. By choice of these points we known that consecutive points lie on curves that differ in their value of I by $\frac{2i_0}{n+1}$. Thus, the learned value of I will differ from the true value of I by maximum of $\frac{4i_0}{n+1}$.

The total number of rounds is upper bounded by $l(2q+1)\lceil \log_2 n \rceil + d$, which is from $2q+1$ rounds for each query and a maximum of $\lceil \log_2 n \rceil$ queries for each of the l line, followed by d dummy rounds. □

Reminder of Theorem 1. *The near-optimal strategy profile (defined above) is an* $(\epsilon_{\mathcal{A},k}, 0)$*-SPE for the game* $G_{\mathcal{A},k}$*, where* $\epsilon_{\mathcal{A},k}$ *is*

$$\max\left(\max_{v^t,a^t}(U^{D+\delta I}_{\max} - U^{D}_{\max}),\ \max_{v^t,a^t}(U^{ND+\delta I}_{\max} - U^{ND}_{\max})\right) + \epsilon_{th}\max_{\alpha\in[0,1]}\left(\sum_{j=0}^{m-1}\delta^j_{\mathcal{D}}E(r(\vec{\mathbf{O}}^t, j))[U_k, U_k, \alpha]\right)$$

Proof. First the easy case for the employee: the employee always plays a best response. When deterred she is indifferent among any a^t, so choice of a^t does not matter in that case. Thus, there is 0 benefit for the employee by deviating with the history-independent strategy followed. There are two terms in the $\epsilon_{\mathcal{A},k}$ bound for the organization. The first term bounds the profit from deviation due to the fact that the true I is not known. The second term further bounds the profit from deviation due to the fact that the distribution D^0_t is unknown.

Note that we have lifted the action space of $\mathcal{D}$ to commitment functions. Thus, we need to compare the given commitment with other commitment functions. First, note that if the regions were known properly, and D^t_0 known then it is possible to find the commitment that is optimal cost for each fixed value of a^t. Then, it is enough to bound the difference in utility of the the audit commitment function to this optimal commitment function across all values of a^t. We perform the analysis for any fixed a^t, then taking the maximum over all a^t to bound the difference in utility when $\mathcal{D}$ could move to the optimal commitment. We first compare the audit commitment to itself when the true regions are known, then assuming true regions are known we compare the audit commitment to the optimal commitment. Then using triangle inequality we get the required difference for a fixed a^t. Then using the fact that $\max_x f(x) + g(x) \leq \max_x f(x) + \max_x g(x)$ we get the required bound for all a^t.

Suppose the near-optimal strategy profile finds a point in the region $R^I_D \backslash R^I_{\delta I}$. The largest true deterred region can be $R^I_D \cup R^I_{\delta I}$. Thus, $U^{D+\delta I}_{\max} - U^{D}_{\max}$ represents the maximum profit the organization could have obtained by deviating to another point using the true deterred region in near-optimal strategy, with some fixed value of v^t and a^t. Then the maximum taken over v^t and a^t gives the maximum possible profit by deviation for the deterred region if the true deterred region were known. Similar argument shows that the maximum profit by deviation for non-deterred region is $\max_{v^t,a^t}(U^{ND+\delta I}_{\max} - U^{ND}_{\max})$. Thus, the absolute maximum profit from deviation for any region is given by the maximum of these two quantities.

Next, assume that the true deterred region R_D is known, so is the non-deterred region R_{ND}. We have already show above that the maximum profit from deviation that the organization would get using near-optimal strategy with R_D instead of $R^I_D \backslash R^I_{\delta I}$, and R_{ND} instead of $R^I_{ND} \backslash R^I_{\delta I}$. Assume that the true regions are known and near-optimal strategy outputs (P, α) to be played by the organization. We use the simplified notation with the game under consideration being $G_{\mathcal{A},k}$. Denote by $f(P, \alpha)$ the function $E(\mathbf{Rew}^t_{\mathcal{D}})[0]$, by $g(P, \alpha)$ the function $E(\mathbf{Rew}^t_{\mathcal{D}})[a^t]$ and by $h(P, \alpha)$ the function $E(\mathbf{Rew}^t_{\mathcal{D}})[D^t_0]$. The function maximized by (P, α) is

$$U_D(P, \alpha) = (1 - \epsilon_{th})f(P, \alpha) + \epsilon_{th}g(P, \alpha)\ ,$$

in R_D and is

$$U_{ND}(P, \alpha) = (1 - \epsilon_{th})g(P, \alpha) + \epsilon_{th}g(P, \alpha)$$

in R_{ND}. Suppose D^t_0 was known and the point (P', α') is obtained by maximizing

$$U'_D(P, \alpha) = (1 - \epsilon_{th})f(P, \alpha) + \epsilon_{th}h(P, \alpha)$$

in the R_D region and

$$U'_{ND}(P, \alpha) = (1 - \epsilon_{th})g(P, \alpha) + \epsilon_{th}h(P, \alpha)$$

in the R_{ND} region. We emphasize that the function U' is the true expected utility. Consider two different cases

- (P, α) and (P', α') both lie in the same region, say R_D. Then, the maximum benefit to be gained out of deviation is $U'_D(P', \alpha') - U'_D(P, \alpha)$, which is

  $$(1 - \epsilon_{th})(f(P', \alpha') - f(P, \alpha)) + \epsilon_{th}(h(P', \alpha') - h(P, \alpha))$$

  Also, since $U_D(P, \alpha) \geq U_D(P', \alpha')$ we have

  $$\epsilon_{th}(g(P, \alpha) - g(P', \alpha')) \geq (1 - \epsilon_{th})(f(P', \alpha') - f(P, \alpha))$$

  Thus, the maximum benefit is upper bounded by

  $$\epsilon_{th} \left(g(P, \alpha) - g(P', \alpha') + h(P', \alpha') - h(P, \alpha)\right) .$$

  The upper bound is same for the non-deterred case, since in that case the function $f(.,.)$ is replaced by $g(.,.)$ in both U and U' and the exact same calculation as above yields the same bound.

- (P, α) and (P', α') both lie in different regions, say R_D and R_{ND} respectively. Then, the maximum benefit to be gained out of deviation is $U'_{ND}(P', \alpha') - U'_D(P, \alpha)$, which is

  $$(1 - \epsilon_{th})(g(P', \alpha') - f(P, \alpha)) + \epsilon_{th}(h(P', \alpha') - h(P, \alpha)) .$$

  Also, since $U_D(P, \alpha) \geq U_{ND}(P', \alpha')$ we have

  $$\epsilon_{th}(g(P, \alpha) - g(P', \alpha')) \geq (1 - \epsilon_{th})(g(P', \alpha') - f(P, \alpha)) .$$

  Thus, the maximum benefit is upper bounded by

  $$\epsilon_{th} \left(g(P, \alpha) - g(P', \alpha') + h(P', \alpha') - h(P, \alpha)\right) .$$

  Now suppose that (P, α) and (P', α') lie in R_{ND} and R_D respectively. Then, the maximum benefit to be gained out of deviation is $U'_D(P', \alpha') - U'_{ND}(P, \alpha)$, which is

  $$(1 - \epsilon_{th})(f(P', \alpha') - g(P, \alpha)) + \epsilon_{th}(h(P', \alpha') - h(P, \alpha)) .$$

  Also, since $U_{ND}(P, \alpha) \geq U_D(P', \alpha')$ we have

  $$\epsilon_{th}(g(P, \alpha) - g(P', \alpha')) \geq (1 - \epsilon_{th})(f(P', \alpha') - g(P, \alpha)) .$$

  Thus, the maximum benefit is upper bounded by

  $$\epsilon_{th} \left(g(P, \alpha) - g(P', \alpha') + h(P', \alpha') - h(P, \alpha)\right) .$$

The above cases show that the upper bound for profit from deviation in one round is always

$$\epsilon_{th} \left(g(P, \alpha) - g(P', \alpha') + h(P', \alpha') - h(P, \alpha)\right) .$$

Using definition of expected rewards we have

$$g(P, \alpha) = -C(\alpha^t a^t) - e(P^t)$$

$$h(P, \alpha) = -C(\alpha^t a^t) - e(P^t) - \sum_{j=0}^{m-1} \delta_o^j E_{D_0^t}(E(r(\mathbf{O}^t, j)))$$

Note that for any P, α

$$h(P, \alpha) - g(P, \alpha) = -\sum_{j=0}^{m-1} \delta_o^j E_{D_0^t}(E(r(\mathbf{O}^t, j))) \leq 0 ,$$

thus, the upper bound above is further bounded by

$$\epsilon_{th}\left(g(P,\alpha)-h(P,\alpha)\right)\ ,$$

which is given by

$$\epsilon_{th}\left(\sum_{j=0}^{m-1}\delta_{\mathcal{D}}^{j}E_{D_0^t}(E(r(\mathbf{O}^t,j)))\right)\ .$$

Observe that the above term is maximized over choice of D_0^t when D_0^t places all probability mass on a^t (for any α), i.e., $v^t = a^t$. Also, the expected value of r should be increasing in v^t (since higher v^t means higher detected violations), and $v^t = a^t$ takes a maximum value of U_k for game $G_{\mathcal{A},k}$. Thus, the above term is upper bounded by

$$\epsilon_{th}\max_{\alpha\in[0,1]}\left(\sum_{j=0}^{m-1}\delta_{\mathcal{D}}^{j}E(r(\mathbf{O}^t,j))[U_k,U_k,\alpha]\right)$$

Now, add the two bounds to get maximum profit from deviation in one round. Further, using Property 1 and noting that the strategy is history independent for $G_{\mathcal{A},k}$ we now obtain the desired result. □

Claim 1. *Assume function instantiations from Section 3.3. Thus, given $\nu(\alpha) = \mu\alpha - (\mu-1)\alpha^2$, we must have $\mu \leq 2$. Further, assuming $C + 2(R_{int} - R_{ext}p) \geq 0$ and $R_{int} \leq R_{ext}p$, the $\epsilon_{\mathcal{A},k}$ from Theorem 1 is given by $\epsilon_{th}U_k\max(R_{int}, R_{ext}p) + \Delta I_{\mathcal{A},k}$, where $\Delta I_{\mathcal{A},k}$ is*

$$2\delta I\min\left(\frac{e}{p},\frac{U_kC}{\mu(1-p)(I-\delta I)}\right)$$

Using values from end of Section 3 we can get $\epsilon_{th}U_k\max(R_{int}, R_{ext}p) = 0.03 * 40 * 150 = 180$, also, the minimum in $\Delta I_{\mathcal{A},k}$ is for $e/p = 20$. Assuming, $\delta I = 0.5$ (remember $i_0 = 1$, assume the learning reduces region of uncertainty by half), we have $\Delta I_{\mathcal{A},k} = 20$. Thus, we get $\epsilon_{\mathcal{A},k} = \$200$.

Proof. Remember that Note that for $\nu(\alpha) \leq 1$ to hold, it must be that $\mu\alpha - (\mu-1)\alpha^2 \leq 1$ for $\alpha \in [0,1]$. It can be readily verified that this happens only when $\mu \leq 2$. Remember the linear functions assumption means $C(s^t) = Cs^t$ and $e(P^t) = eP^t$.

$$\max_{\alpha\in[0,1]}\sum_{j=0}^{m-1}\delta_{\mathcal{D}}^{j}E(r(\mathbf{O}^t,j))[U_k,U_k,\alpha] = U_kR_{ext}p + U_k\max_{\alpha\in[0,1]}(R_{int}-R_{ext}p)\nu(\alpha)$$

The relevant part to maximize can be expanded as

$$(R_{int}-R_{ext}p)(\mu\alpha-(\mu-1)\alpha^2)$$

For $\mu < 2$, $(\mu\alpha - (\mu-1)\alpha^2)$ increases with $\alpha \in [0,1]$ (derivative is positive). Thus, if $R_{int} > R_{ext}p$ then $\alpha = 1$ is the maximizer else $\alpha = 0$ is the maximizer. Then, it is not difficult to conclude that the maximum value is $U_k\max(R_{int}, R_{ext}p)$.

Now observe that, since $\mu < 2$, $\mu \geq \mu(\alpha) \geq 1$ The utility function maximized by the organization given the linear function and the example reputation function is (using simple notation)

$$\begin{aligned}-\epsilon_{th}R_{ext}pa^t - eP^t - \alpha^ta^tC - \nu(\alpha^t)a^t\epsilon_{th}(R_{int}-R_{ext}p)\ ,\\ -R_{ext}pa^t - eP^t - \alpha^ta^tC - \nu(\alpha^t)a^t(R_{int}-R_{ext}p)\end{aligned}$$

in the reduced deterred region and reduced non-deterred region respectively. Observe that for the non-deterred case, using assumption $C + 2(R_{int} - R_{ext}p) \geq 0$ implies $C + \mu(\alpha)(R_{int} - R_{ext}p) \geq 0$, since $2 > \mu \geq \mu(\alpha) \geq 1$ and all

quantities C, R_{int} and $R_{ext}p$ are positive. Thus, the maximizer in non-deterred region is always $0, 0$, irrespective of the value of I, hence the difference in costs is zero for the cases when I is known perfectly and when there is an error δI.

Assume $U^{D+\delta I}_{\max}$ occurs for a point P', α' and $U^{D}_{\max}$ happens for a point P, α, and learned valued of personal benefit is I. The interesting case is when $P', \alpha' \neq P, \alpha$ and P, α lies on the curve defined by $I + \delta I$. Then suppose P', α' lie on the curve defined by $I + \delta I - \zeta$ for $2\delta I \geq \zeta \geq 0$. Suppose P', α'' and P'', α' are points on the curve defined by $I + \delta I$ obtained by drawing straight lines from the point P', α'. Thus, $P' \leq P''$ and $\alpha' \leq \alpha''$. Note that since $P'(\nu(\alpha') + p(1 - \nu(\alpha'))) = I + \delta I - \zeta$ and $\nu(\alpha), p \leq 1$, we can claim that $P' \geq I - \delta I$. Then we have

$$\zeta = P''(\nu(\alpha') + p(1 - \nu(\alpha'))) - P'(\nu(\alpha') + p(1 - \nu(\alpha')))$$

or

$$P'' - P' = \frac{\zeta}{\nu(\alpha')(1-p) + p} \leq \frac{\zeta}{p}$$

Also,

$$\zeta = P'(\nu(\alpha'') + p(1 - \nu(\alpha''))) - P'(\nu(\alpha') + p(1 - \nu(\alpha')))$$

or

$$\nu(\alpha'') - \nu(\alpha') = \frac{\zeta}{(1-p)P'} \leq \frac{\zeta}{(1-p)(I - \delta I)}$$

Note that

$$\nu(\alpha'') - \nu(\alpha') = \mu(\alpha'' - \alpha') - (\mu - 1)(\alpha'' - \alpha')((\alpha'' + \alpha'))$$

thus, $\nu(\alpha'') - \nu(\alpha') > \mu(\alpha'' - \alpha')$ and hence

$$(\alpha'' - \alpha') \leq \frac{\zeta}{\mu(1-p)(I - \delta I)}$$

Also, $U^D(P, \alpha) > U^D(P'', \alpha')$ and $U^D(P, \alpha) > U^D(P', \alpha'')$ and, $U^{D+\delta I}_{\max} - U^{D}_{\max} = U^{D+\delta I}(P', \alpha') - U^D(P, \alpha)$ means that

$$U^{D+\delta I}_{\max} - U^{D}_{\max} \leq \min(U^{D+\delta I}(P', \alpha') - U^D(P'', \alpha'), \\ U^{D+\delta I}(P', \alpha') - U^D(P', \alpha''))$$

Also, $U^{D+\delta I}(P', \alpha') - U^D(P'', \alpha')$ is given by

$$-e(P' - P'') \leq \frac{e\zeta}{p}$$

Also, $U^{D+\delta I}(P', \alpha') - U^D(P', \alpha'')$ is given by

$$-a^t(\alpha' - \alpha'')C - a^t(\nu(\alpha') - \nu(\alpha''))\epsilon_{th}(R_{int} - R_{ext}p)$$

which can be simplified to

$$a^t(\alpha'' - \alpha')(C + (\mu - (\mu - 1)(\alpha'' + \alpha'))\epsilon_{th}(R_{int} - R_{ext}p))$$

Using result $1 \leq \mu \leq 2$, we have $2 \geq \mu - (\mu - 1)(\alpha + \alpha') \geq 0$. Using assumption $C + 2(R_{int} - R_{ext}p) \geq 0$ we can say that $(C + (\mu - (\mu - 1)(\alpha + \alpha'))\epsilon_{th}(R_{int} - R_{ext}p)) \geq 0$. Also, since $R_{int} \leq R_{ext}p$, we have $(C + (\mu - (\mu - 1)(\alpha + \alpha'))\epsilon_{th}(R_{int} - R_{ext}p)) \leq C$. Thus, using the inequalities above and $\zeta \leq 2\delta I$, $U^{D+\delta I}_{\max} - U^{D}_{\max}$ is less than

$$2\delta I \min\left(\frac{e}{p}, \frac{a^t C}{\mu(1-p)(I - \delta I)}\right)$$

which is maximized for $a^t = U_k$. □

B.1 Repeated Product Game - Definition and Results

If two players play multiple (repeated) *independent* games in parallel then it is possible to consider a composition of these games which is itself a (repeated) game. By *independent* games we mean that these games are played without any influence from the other games in parallel. We define the composition below for a repeated game.

Definition 2. *(Repeated Product Games) Let the two players play the independent one-shot stage games $G1, G2..., Gn$ in parallel in each round of the corresponding n repeated games. A composition of the n stage games is a single-shot game G given by player i's ($i = 1, 2$) action space $S_i = S1_i \times S2_i ... \times Sn_i$, and the payoff function $r_i(s_1, s_2) = \sum_{j=1}^{n} r_i(sj_1, sj_2)$ where $sj_i \in Sj_i$ and $s_i \in S_i$. A repeated product game is a repeated game with the stage game in every round given by G.*

We can extend the above definition to games with imperfect monitoring and public signaling, similar to the manner in which a standard repeated game is extended. Observe that any strategy σ of a repeated product game can be decomposed into strategies $\sigma 1, ..., \sigma n$ of the component games, because of the independence assumption of the component games. This decomposition leads to the following useful results summarized in the lemma below:

Lemma 3. *Let RG be a repeated product game with the stage game given by G, such that G is a parallel composition of $G1, G2..., Gn$ as defined in Definition 2. Consider a strategy σ of RG with the decomposition into strategy σi for each component game. Then*

- *The strategy σ is a SPE iff the strategy σi is a SPE for the repeated game with stage game Gi for all i.*
- *The strategy σ is an $(\sum_{i=1}^{n} \epsilon_1 i, \sum_{i=1}^{n} \epsilon_2 i)$-SPE if the strategy σi is a $(\epsilon_1 i, \epsilon_2 i)$-SPE for the repeated game with stage game Gi for all i.*

Proof. For the first case assume σi is a SPE for the repeated game with stage game Gi for all i. Since σ is given by $\sigma 1, ..., \sigma n$ any unilateral deviation from σ results in a unilateral deviation from one or more of $\sigma 1, ..., \sigma n$, suppose it is σj. By assumption that is not profitable for repeated game given by the stage game Gj. Since the payoff in G is the sum of payoffs in $G1, ..., Gn$ and payoffs of games other than j^{th} game remains same, the deviation is not profitable for G also.

The other direction is very similar. Assume σ is a SPE. Since σ is given by $\sigma 1, ..., \sigma n$ any unilateral deviation from σj results in a unilateral deviation from σ. By assumption that is not profitable for repeated game given by the stage game G. Since the payoff in G is the sum of payoffs in $G1, ..., Gn$ and payoffs of games other than j^{th} game remains same, the deviation is not profitable for Gj also.

Next, for the second part since the payoff of G is the sum of payoff's of Gi's and any any unilateral deviation from σ results in a unilateral deviation from one or more of $\sigma 1, ..., \sigma n$, then it is not difficult to check that the profit from deviation will not more than the sum of profit from deviation in each of the repeated games defined by $G1, ..., Gn$. Thus, the maximum profit from deviation for player j is $\sum \epsilon_j i$. □

B.2 Determining P_f and Punishment

In addition to the action dependent utilities above, the players also get an fixed utility in each round of $\mathcal{G}_{\mathcal{A}}$, which is the salary $Sal_{\mathcal{A}}$ for $\mathcal{A}$ and the value created by the employee $g \times Sal_{\mathcal{A}}$ for $\mathcal{D}$. Note that this is the salary and value created for the duration of one audit cycle. Also, note that this fixed utility is not part of any game $\mathcal{G}_{\mathcal{A},k}$. Let R_k be the maximum loss of reputation possible for violation of type k. We assume that the maximum punishment $\vec{P}_{fire,\mathcal{A}}(k)$ rate for each type k is proportional to R_k. Since the employee can make mistakes, in the worst case he can lose an expected amount of $\epsilon_{th} \sum_k \vec{P}_{f,\mathcal{A}}(k) U_k$. This loss must be less than a fixed fraction net of $Sal_{\mathcal{A}}$, or else the employee is better off quitting and getting betters expected payoff in every round in some other job. Thus, we must have $\epsilon_{th} \sum_k \vec{P}_{f,\mathcal{A}}(k) U_k = net \cdot Sal_{\mathcal{A}}$, which yields a value $\vec{P}_{f,\mathcal{A}}(k) = R_k net \cdot Sal_{\mathcal{A}} / (\epsilon_{th} \sum_k R_k U_k)$. Observe that an employee with higher salary can be punished more. For example, suppose $\mathcal{A}$ does three types k, k' of tasks such that in every week $U_k = 40, U_{k'} = 10$ and $R_{k'} = 12 R_k$ and $net = 0.1$ with weekly salary \$500. Then, $\vec{P}_f(k) = 10.4$.

Next, consider the case the the employee is non-deterred for violations of type k. Then suppose the expected loss to the organization in every round for such a case is maximum of $U_k L_k$, where L_k is maximum per violation cost

(dependent on α) that can be calculated from our model. In such a case if it happens that $U_K L \geq (g-1)Sal_{\mathcal{A}}$ then the organization obtains no benefit from employing $\mathcal{A}$. Thus, in such a case the organization must fire the employee.

B.3 Budget Optimization Problem

If the overall budget in an audit cycle is given by B then we must have $\sum_{\mathcal{A},k} \vec{b}^t_{\mathcal{A}}(k) \leq B$. Further let $f(\vec{b}^t_{\mathcal{A}}(k))$ denote the expected utility in game $\mathcal{G}_{\mathcal{A},k}$ from the equilibrium computed in Section 4. Note that the maximum of the cost functions in deterred and non-deterred regions are continuous in $\vec{b}^t_{\mathcal{A}}(k)$, since the regions themselves change continuously with change in $\vec{b}^t_{\mathcal{A}}(k)$. Since the equilibrium utility involves taking maximum of two continuous functions, using fact that max of two functions is continuous, we get that $f(\vec{b}^t_{\mathcal{A}}(k))$ is continuous. Thus, the optimal allocation of budget is to solve the following non-linear optimization problem

$$\text{maximize } \sum_{\mathcal{A},k} f(\vec{b}^t_{\mathcal{A}}(k)) \text{ subject to } \sum_{\mathcal{A},k} \vec{b}^t_{\mathcal{A}}(k) \leq B$$